```
            GenCore version 5.1.6
        Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on:   February  9, 2005, 17:51:02 ; Search time 229 Seconds
                                        (without alignments)
                             3741.085 Million cell updates/sec

Title:       US-09-889-874A-23
Perfect score: 8879
Sequence:    1 VYIKFLKLFRRITMSDNNEF.........PRKIILGRTEKTVKPKTFRP 1673

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:    1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters:   1612378

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   UniProt 03:*
             1: uniprot_sprot:*
             2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	8815	99.3	1660	2	Q9EVR7	Q9evr7 xenorhabdus
2	1632	18.4	1625	2	Q7MB38	Q7mb38 photorhabdu
3	1614	18.2	1562	2	Q883V9	Q883v9 pseudomonas
4	1598	18.0	1385	2	Q7NWU2	Q7nwu2 chromobacte
5	1583	17.8	1775	2	Q7N4A7	Q7n4a7 photorhabdu
6	1553	17.5	773	2	Q9S6J1	Q9s6j1 coxiella bu
7	1530.5	17.2	774	2	O52880	O52880 coxiella bu
8	1416	15.9	1632	2	Q88LP7	Q88lp7 pseudomonas
9	1214	13.7	709	2	Q45948	Q45948 coxiella bu
10	1037	11.7	1290	2	Q88LP2	Q88lp2 pseudomonas
11	932.5	10.5	982	2	Q883W6	Q883w6 pseudomonas
12	920.5	10.4	1669	2	Q87VG6	Q87vg6 pseudomonas
13	919.5	10.4	505	2	Q45905	Q45905 coxiella bu
14	919.5	10.4	526	2	Q9X626	Q9x626 coxiella bu
15	916.5	10.3	526	2	O52883	O52883 coxiella bu
16	912.5	10.3	528	2	Q45949	Q45949 coxiella bu
17	884	10.0	589	2	Q45902	Q45902 coxiella bu
18	648	7.3	762	2	Q93IJ8	Q93ij8 vibrio pomm
19	552	6.2	370	2	Q7NVR8	Q7nvr8 chromobacte
20	486	5.5	389	2	Q7NVA0	Q7nva0 chromobacte
21	478.5	5.4	310	2	Q45904	Q45904 coxiella bu
22	450.5	5.1	2217	2	Q8TP72	Q8tp72 methanosarc
23	450	5.1	380	2	Q7NW05	Q7nw05 chromobacte
24	439	4.9	2554	2	Q72QR5	Q72qr5 leptospira
25	432	4.9	2334	1	WAPA_BACSU	Q07833 bacillus su
26	429.5	4.8	336	2	Q45946	Q45946 coxiella bu
27	420.5	4.7	2364	2	Q82R58	Q82r58 streptomyce
28	418.5	4.7	1835	2	Q6MD86	Q6md86 parachlamyd
29	414.5	4.7	2221	2	Q73BZ3	Q73bz3 bacillus ce
30	396	4.5	1513	2	Q7NY44	Q7ny44 chromobacte
31	395	4.4	820	2	Q82QR1	Q82qr1 streptomyce
32	395	4.4	2246	2	Q73C66	Q73c66 bacillus ce
33	393.5	4.4	1976	2	Q8D4R5	Q8d4r5 vibrio vuln
34	392	4.4	1250	2	Q82RX1	Q82rx1 streptomyce
35	391.5	4.4	2370	2	Q82RE3	Q82re3 streptomyce
36	388.5	4.4	1826	2	Q98Z27	Q98z27 rhizobium l
37	386	4.3	2167	2	Q92EK5	Q92ek5 listeria in
38	386	4.3	2224	2	Q81U00	Q81u00 bacillus an
39	382.5	4.3	1117	2	Q8CK70	Q8ck70 streptomyce
40	381	4.3	2183	2	O85585	O85585 streptomyce
41	376.5	4.2	2082	2	Q9ACP4	Q9acp4 streptomyce
42	374.5	4.2	1511	2	Q6MF99	Q6mf99 parachlamyd
43	371	4.2	1473	2	Q6JCV7	Q6jcv7 xanthomonas
44	368.5	4.2	1427	2	Q6CZ75	Q6cz75 erwinia car
45	368	4.1	2031	2	Q63XE6	Q63xe6 burkholderi

ALIGNMENTS

```
RESULT 1
Q9EVR7
ID   Q9EVR7       PRELIMINARY;      PRT;   1660 AA.
AC   Q9EVR7;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Nematicidal protein 2.
GN   Name=xnp2;
OS   Xenorhabdus bovienii.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Xenorhabdus.
OX   NCBI_TaxID=40576;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=I73;
RA   Morgan J.A.W., Ousley M., Ellis D., Jarrett P.;
RL   Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ296651; CAC19493.1; -.
DR   InterPro; IPR006530; YD.
DR   Pfam; PF05593; RHS_repeat; 5.
DR   TIGRFAMs; TIGR01643; YD_repeat_2x; 4.
DR   TIGRFAMs; TIGR01643; RHS_repeat; 5.
SQ   SEQUENCE   1660 AA;  185738 MW;  AFABEA20AD70B164 CRC64;

Query Match             99.3%;  Score 8815;  DB 2;  Length 1660;
Best Local Similarity   100.0%;
Matches 1660;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy     14 MSDNNEFFTQANNFTSAVSGGVDPRTGLYNIQITLGHIVGNGNLGPTLPLTLSYSPLNKT 73
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MSDNNEFFTQANNFTSAVSGGVDPRTGLYNIQITLGHIVGNGNLGPTLPLTLSYSPLNKT 60

Qy     74 DIGFGIGFNFGLSVYDRKNSLLSLSTGENYKVIETDKTVKLQQKKLDNLRFEKDLKENCY 133
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 DIGFGIGFNFGLSVYDRKNSLLSLSTGENYKVIETDKTVKLQQKKLDNLRFEKDLKENCY 120

Qy    134 RIIHKSGDIEVLTGFNNNAFDLKVPKKLLNPAGHAIYIDWNFEATQPRLNRIYDDLDGHD 193
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 RIIHKSGDIEVLTGFNNNAFDLKVPKKLLNPAGHAIYIDWNFEATQPRLNRIYDDLDGHD 180

Qy    194 IPLLNLEYQGLIKTILTLFPGQKEGYRTELRFLNRQLNSIHNFSLGNENPLTWSFGYTPI 253
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 IPLLNLEYQGLIKTILTLFPGQKEGYRTELRFLNRQLNSIHNFSLGNENPLTWSFGYTPI 240

Qy    254 GKNGILGQWITSMTAPGGLKETVNYSNNNQGHHFPQSANLPVLPYVTLMKQVPGAGQPAI 313
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GKNGILGQWITSMTAPGGLKETVNYSNNNQGHHFPQSANLPVLPYVTLMKQVPGAGQPAI 300

Qy    314 QAEYSYTSHNYVGGGSNGIWNNKLDNLYGLMTEYNYGSTESRRYKDKEGHDQIVRIERTY 373
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 QAEYSYTSHNYVGGGSNGIWNNKLDNLYGLMTEYNYGSTESRRYKDKEGHDQIVRIERTY 360

Qy    374 NNYHLLTSECKQQNGYIQTTETAYYAIIGHNFDSQPSQFQLPKTKTETWRSADNSYRSEI 433
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db

Qy   361 NNVHLLTSECKQONGYIQTETAYYAIIGHNPDSQPSFQOLPKTKTETWRSADNSYRSEI 420

Qy   434 TETTPFDESGNPLTKVIKDKKTQKIISPSTHWEYYPPAGEVDNCPPEPYGTRFVKKIIQT 493
             |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 TETTFFDESGNPLTKVIKDKKTQKIISPSTHWEYYPPAGEVDNCPPEPYGTRFVKKIIQT 480

Qy   494 PYDSEFKDDPEKFIQYRYSLIGSQSHVTLKIEERHYSATQLLNSTLFQYNTDKSELGRLL 553
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 PYDSEFKDDPEKFIQYRYSLIGSQSHVTLKIEERHYSATQLLNSTLFQYNTDKSELGRLL 540

Qy   554 KQTECTKGENGKTYSVVHKFYTKQDDTLQQSHSITTHDNFTIHRSQVRSRYTGRLFSDT 613
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 KQTECTKGENGKTYSVVHKFYTKQDDTLQQSHSITTHDNFTIHRSQVRSRYTGRLFSGRLL 600

Qy   614 DTKDIVTQMSYDKLGRLLTRTLNSGTPYANTLTYDYELNNLQDDNRPPFVITTTDVNGNQ 673
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 DTKDIVTQMSYDKLGRLLTRTLNSGTPYANTLTYDYELNNLQDDNRPPFVITTTDVNGNQ 660

Qy   674 LRNEFPDGAGRHVSQCLKDSDGDGKFYTIHTQQYDEQGRHHTSTYSDYLTNGRQQTDPDKV 733
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   661 LRNEFPDGAGRHVSQCLKDSDGDGKFYTIHTQQYDEQGRHHTSTYSDYLTNGRQQTDPDKV 720

Qy   734 HLSMSKSYDNWGQIANTHWSYGVSEKITVDPITLTATKQLQSNSNNVQTGKEVTTYTPSQ 793
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   721 HLSMSKSYDNWGQIANTHWSYGVSEKITVDPITLTATKQLQSNSNNVQTGKEVTTYTPSQ 780

Qy   794 QPIQITLFDEAGHLQSCHTLTRDGWDRVRKETDAIGQCTTYQYDNYNRVIQITLPDGTIV 853
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   781 QPIQITLFDEAGHLQSCHTLTRDGWDRVRKETDAIGQCTTYQYDNYNRVIQITLPDGTIV 840

Qy   854 NRKYAPFSTDTLITDIRVNGISLGQOTFDGLSRLTQSQDGGRVWAYTYSAGNDQCPSTVI 913
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   841 NRKYAPFSTDTLITDIRVNGISLGQOTFDGLSRLTQSQDGGRVWAYTYSAGNDQCPSTVI 900

Qy   914 TPDGQFIHYQYQPELDDAVLQVASNEITQQFSYNPVTGALLKAVAEGQSLTPIYYPSGRL 973
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   901 TPDGQFIHYQYQPELDDAVLQVASNEITQQFSYNPVTGALLKAVAEGQSLTPIYYPSGRL 960

Qy   974 KMENINDMKKMSYLWTLRGLENGYTDLTGTIQKISRDTHGRVTQIKDSSIKTLNYDDLN 1033
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   961 KMENINDMKKMSYLWTLRGLENGYTDLTGTIQKISRDTHGRVTQIKDSSIKTLNYDDLN 1020

Qy  1034 RHIGSQVTDLATGHMLTTVEFDGLNREIGRKLCDSSGHTLDIQQSWLKTQQLANRIVKL 1093
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1021 RHIGSQVTDLATGHMLTTVEFDGLNREIGRKLCDSSGHTLDIQQSWLKTQQLANRIVKL 1080

Qy  1094 NGVLQRRTEQYSYDSRNRLNQYKCDGAECPTDKYGHSIVTQNFTYDIYGNITACHTTFADG 1153
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1081 NGVLQRRTEQYSYDSRNRLNQYKCDGAECPTDKYGHSIVTQNFTYDIYGNITACHTTFADG 1140

Qy  1154 TEDHATFKFANPTDPCQLTEVHHTHPDMPDNIRLKYDKAGRVINITDNHGNTENFTYDTL 1213
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1141 TEDHATFKFANPTDPCQLTEVHHTHPDMPDNIRLKYDKAGRVINITDNHGNTENFTYDTL 1200

Qy  1214 GRLQNGQGSVYGYDPLNRIVSQKTDTLDCELYRETMLVNEVRNGEMIRLLRTGETIIAQ 1273
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1201 GRLQNGQGSVYGYDPLNRIVSQKTDTLDCELYRETMLVNEVRNGEMIRLLRTGETIIAQ 1260

Qy  1274 QRASKVLLTGTDSQQSVILTSDKQNLSQEAYSAYGKHKSTANDASILGYNGERADPVSGV 1333
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1261 QRASKVLLTGTDSQQSVILTSDKQNLSQEAYSAYGKHKSTANDASILGYNGERADPVSGV 1320

Qy  1334 THLGNGYRSYDPTLMRFHTPDSLSPFGAGGINPYSYCLGDPINRSDPSGHLSWQAWTGIG 1393
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1321 THLGNGYRSYDPTLMRFHTPDSLSPFGAGGINPYSYCLGDPINRSDPSGHLSWQAWTGIG 1380

Qy  1394 MGIAGLLLTIATGMAIAAAGGIAAAIASTSTTALAFGALSVTSDITSIVSGALEDASPK 1453
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1381 MGIAGLLLTIATGMAIAAAGGIAAAIASTSTTALAFGALSVTSDITSIVSGALEDASPK 1440

Qy  1454 ASSILGWVSMGMGAAGLAESAIKGGTKLATHLGAFAEDGENALLKSTSESSRIKWGVTRS 1513
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1441 ASSILGWVSMGMGAAGLAESAIKGGTKLATHLGAFAEDGENALLKSTSESSRIKWGVTRS 1500
```

```
Qy  1514 LDREIVRNEEGQVIKDHSRGYTDNFMGKGEQAILVHGDKDGFLYHTEGNKHNGKGPYTRH 1573
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1501 LDREIVRNEEGQVIKDHSRGYTDNFMGKGEQAILVHGDKDGFLYHTEGNKHNGKGPYTRH 1560

Qy  1574 TPEQLVDYLKDNNIVDLTQGGDKPVHLLSCYGKSSGAADKMAKYINRPVIAYSNKPTISQ 1633
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1561 TPEQLVDYLKDNNIVDLTQGGDKPVHLLSCYGKSSGAADKMAKYINRPVIAYSNKPTISQ 1620

Qy  1634 GLARIERKDFFLKSTVHSYDPRKIILGRTEKTVKPKTFRP 1673
             ||||||||||||||| |||||||||||||||||||||||
Db  1621 GLARIERKDFFLKSTYHSYDPRKIILGRTEKTVKPKTFRP 1660
```

```
RESULT 2
Q7MB38
ID   Q7MB38                PRELIMINARY;       PRT;   1625 AA.
AC   Q7MB38;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Similar to the nematicidal protein 2. Probable membrane protein.
GN   OrderedLocusNames=plu2222;
OS   Photorhabdus luminescens (subsp. laumondii).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Photorhabdus.
OX   NCBI_TaxID=141679;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TT01;
RX   MEDLINE=22957627; PubMed=14528314;
RA   Duchaud E., Rusniok C., Frangeul L., Buchrieser C., Givaudan A.,
RA   Taourit S., Bocs S., Boursaux-Eude C., Chandler M., Charles J.-F.,
RA   Dassa E., Derose R., Derzelle S., Freyssinet G., Gaudriault S.,
RA   Medigue C., Lanois A., Powell K., Siguier P., Vincent R., Wingate V.,
RA   Zouine M., Glaser P., Boemare N., Danchin A., Kunst F.;
RT   "The genome sequence of the entomopathogenic bacterium Photorhabdus
RT   luminescens.";
RL   Nat. Biotechnol. 21:1307-1313(2003).
DR   EMBL; BX571866; CAE14515.1; -.
DR   PhotoList; plu2222; -.
DR   InterPro; IPR006530; YD.
DR   Pfam; PF05593; RHS_repeat; 7.
DR   TIGRFAMs; TIGR01643; YD_repeat_2x; 3.
KW   Complete proteome.
SQ   SEQUENCE   1625 AA;   180118 MW;   BFA362F4ECB3EA97 CRC64;

Query Match             18.4%;   Score 1632;   DB 2;   Length 1625;
Best Local Similarity   31.0%;   Pred. No. 1.1e-77;
Matches  494;  Conservative  245;  Mismatches  653;  Indels  204;  Gaps   59;

Qy    18 NEFFTQANNFTSAVSGGVDPRTGLYNIQITLGHIVGN--GNLGPTLPLTLSYSPLNKTDI 75
          : :: .:|  ||  :.:| .|.  ||.||.||.:   |:| |  : :
Db     4 NDIYSNAFNFGSYINTGVDPRTGQYSANINITLRPNNVGNIEQV--LNLSFSPLTTLNN 61

Qy    76 GFGIGFNFGLSVYDRKNSLLSLSTGENYK---VIETDKTVKLQQKKLDNLRFEKDLKENC 132
          ||.|| .|.|: |:.:    |.  :.|.||     .|:|:|.. .:  |.|  :: .|
Db    62 GFGIGWRFSLTMLDVKTLTFSRSNGEQFKCKLPPNNNDISFKDKKLKDLRVK--LDSNT 120

Qy   133 YRIIHKSGDIEVLTGFNNNAFDLKVPKKLLNPAGHAIYIDWNFEATQPRLNRIYDLDLGH 192
          :.:||.||:: || ::      :.|   |.:|    |.| :    |          ::
Db   121 FYYVNKNGIIETLKRIGSS--DIAKTVALEFPDG----------------EVFDLIYNS 161

Qy   193 DIPLLNLEYQGLIKTILTL-FPGQKEGYRTELRFL-NRQLNSIHNFSLGNENPLTWSFGY 250
          .:: :: : :..||| |.| ||.||:|| |.|..| ||.:|.|   :.:
Db   162 RFALSEIKYRMTGKTYLKLNYSGNN---CTSVEYPDDNNISAKIAFDYRNDYLITVTVPY 218

Qy   251 TPIGK-------------NGIIGQWITSMTAPGGLKETVNYSNNNQGHHFPQSANLPVLP 297
          .. .|             .||.| .:.: |.: |   |:.| :|  |:  :     :| .
Db   219 DASGPIDSARFKMTYQTLKGIF-PVISAFRTPTGYVELVSYKEN--GHKVTDTESIPYAA 275

Qy   298 YVTLMKQVPGAGQPAIQAEYSYTS-HNYVGGGSNGI-WNNKLDNLYGLMTEYNYGSTESR 355
          ::: .
Db   298
```

```
Db   276 ALTIQ--PGNGQPAISKSYEYSVHNFLGYSSGRTSFDSSQDNLYLVTGKYTYSSIE--  330
Qy   356 RYKDKEGHDQIVRIERTYNNYHLLTSECKQQNGYIQTTETAYYAIIGHNFDSQPSQFQLP  415
         :  :   ::  :. ::|||||  :::|:: .:
Db   331 --RVLNGQNVISVTERVFDKFHLMTKEAKTQDNKRIITEITYNEDPSKSFSBQPENLQQP  388

Qy   416 K-------TKTETWRSADNSYRSBITETTFPDESGNPLTKVIKDKKTQKIISPSTHWEYPP  469
                 |: |  ::|    ||||: :|::|:  ||||
Db   389 SHVLTRYTDLQT-----NTSREESVNIKSDDWGNTLL-TETSGIQK------EYVYYPV  436

Qy   470 AGEVDNCPEPYGFTRFVKKIIQTPYDSEFKDDPEKFIQYRYSLIGSQSHVTLKIEERHY  529
         .|| :|| | ||:||:||.  ||    .|| |: :||:|:|:::|| ||::|||||:|
Db   437 NGEGNNCPADPLGFSRFLKSVTQKGSPDAAQSVANRVTSYTYQKLPTFTGAYVK--EVVS  494

Qy   530 SATQLLNSTL---FQY--NTDKSELGRLLKQTECTKGENGKTYSVVHKFTYTKQDDTLQ  583
         .| :: :.:     ||     |||| |: ||      ::   |: : |||  ::::
Db   495 KASETIDSKIVRFTNFYVNSPTNKSH-GSLAKITSVMNNQ----QTVTTFKYEYSDSEMT  548

Qy   584 QSHSITTHDNFTIHRSQVRSRYTGRLFSDTDTKDIVTQMSYDKLGRLLTRTLNSGTPYAN  643
          :..:|: :  . : :.:  .|:|::  :.| :  : |:|:|:.|:::  :::||| :|
Db   549 TNSTVTGPDGTHMESKNVTSIYTHRQLRKVDVNHVITDQSYDLSGRIIGQIIIDPGTTKEI  608

Qy   644 TLTYDYELNNLQDDNRPPFVITTDVNGNQLRNEFDGAGRHVSQCLKDSDG--------D  695
          | | | | :  : ::.. |:: :: || :| . :| |||: .  :|:|:         |
Db   609 KRSYIYQYPGGDENDFWP-VMIEIDSQGIRRKTHYDGMGRICSIEEQDDDGVWGTSGIYQ  667

Qy   696 GKFYTIHTQQYDEQGRHHTSTYSDYLTN----GRQQTDPDKVHLSMSKSYDNWGQIANT  750
         |  . : .:|||:.||:: |:|.|: |     ||:|::|||| |::   |||:||:: :|
Db   668 GTYRKVLARQYDVGLQLVKEISNDWLWDLSANPLTRLTTP--LVTTKTYQYDGWGNRYST  725

Qy   751 HWSYGVSEKITVDPITLATK------QLQSNSNNVQTGKEVTYTYTPSQQPIQITLFDBA  804
          .| | .|:||  |||   |       |:.:.:|| ||| |:|:||| :: ::  : :
Db   726 EYSDGRIELEIHDPITRTITQGVKGLGMLNIQQNNF--------EQPASIKVVVPD  773

Qy   805 GHLQSCHTLTRDGWDRVRKETDAIGGQCTIYQYDNYNRVIQITLPDGTIVNRKYAPFSTDT  864
         | : | || ||||.||.:||.|:  ||  :|:|| : ||::|||| |::.. :| :|:. |
Db   774 GAIYSTRTYRYDGFGRTVTETDAEGYATQIEYDLFDRIVKKTLPDRTILESAYIASFSHEE  833

Qy   865 LITDIRVNGISLGQQTFDGLSRLTQSODGRVWAYTYSAGNDQCPSTVTTP--DGQFIHY  922
         ||: ::|||  |.. |  |||| ||::.|||::: |||.|:||  | :|  ||||.:.|
Db   834 LISALNVNGTQLAGSLVYDGLGRVTRDTVGGRKTEYLYGSQGDK-PIQSVTPAHNKQNIDY  892

Qy   923 QYQPELDDAVLQVASNEITQQFSYNPVTGALLKAVAEGQSLTPI-YYPSGRLKMENIN-D  980
          |  |.: :.| :.|:|:||:|||:  |||||:| :||:|  :  |:||| | :|:. |
Db   893 LY--ALGSVMSKFTETSQQNFSYYQKTGALLSA-TEGVSQSNYSYFPSGVLQHESFSRD  949

Qy   981 MKKMS----YLWTLRGLENGYTDLIGTIQKISRDTHGRVTQIKDSSIKTLNYDDLNRHIG 1037
         .| :|     | :|.||:: : |  .::::|::|..|.:|:.|:|. | | |||:.|:| 
Db   950 NKPISSGDYRYTMSGLIQRHKDSFAHDHVYSYDAEGRLVKTEQSQYATFEYDNVGRLLIT 1009

Qy  1038 SQVTDLATGHMLTTVBFPDGLNREIGRKLCDSSGHTLDIQQSWLKTQQLANRIVKLNGVL 1097
           ||: |:|:|| | ||.|.::| |:| ::  :   ||  : ::| : ::.: : : :|:
Db  1010 TTTKDTTSLSLQLATKIEYDVFDREIKRSLLSDFSIQV-ITLSYTKNNQISQRITSIDGVV 1068

Qy  1098 QRTEQYSYDSRNRLNQYKCDGAECPTDKYGHSIVTQNFTYDIYGNITACHTTFPADGTEDH 1157
         ::.|:|.|||.:|| ||:|:| | |.|::|:.:|:||: ||:|||| :  ::|| ||.:::
Db  1069 MKNERYQYDSNQRLSQYQCEGEQSPVDHTGRVLSQOIHYDQWGNIKRLDNTYRDGKET- 1127

Qy  1158 ATFKFANPTDPCQLTEVHTHPDMPDNIRLKYDKAGRVINIT-DNHGNTENFTYDTLGRL 1216
         . : :::|.|  :| :: :  ::. :|.  ||.|  ::.|    : .: :: ||:..|:
Db  1128 VDYHFSQ-ADPTQLIRI----TSDKQQIELSYDANG---NLTRDEKGQT--LIIYDQNNRL 1177

Qy  1217 ---QNGGGSV---YGYDPLNRLVSQ--KTDTLDCELYYRETMLVNEVRNGEMIRLL---- 1264
             :.||.|    ||||.|||| ||   :|: :.||: |. ||.|: |||| |:|:
Db  1178 VQVKDSKGNLVCQYQYDALNKLTAQVLANGTVNRQ-YYASGNVANVQLGDETITWLSSDK 1236

Qy  1265 --------RTGETIIAQQRASKVLLTGTDSQQSVILTSDKQNLSQEAYSAYGKHKSTAND 1316
                 :||| | ||:.:||.::.:| .||:| :: :  .:: :  || | :: | :
Db  1237 QRLGHQSTRNGESVYYQY-------GTDHNSTVIASQNENELMALSYTPYGFRSLI--- 1285

Qy  1317 ASILGYNGERADPVSGVTHLGNGYRSYDPTLMRFHTPDSLSPFGAGGINPYSYCLGDPIN 1376
         :| | .||  .|||:|. .:|||||.: :|:||||:|||| |||:||:||| |||:|||||
Db  1286 SSLLPGLNGAQVDPVTGWYFIGNGYRVFENPVLMRFHSPDSWSPFGRGGVNPYTYCQGDPIN 1345
```

```
Qy  1377 RSDPSGHLSWQAWTGIGMGIAGLLTIAT--GMAIA----AAGGIAAAIASTSTTALAF 1430
         |  | |||| .: ||||||||| |::|:    | |       |||:||||||| ::|:
Db  1346 RIDLNGHLSAGGILGIVLGAIGIIVGVVSLGAGAAISAGLIAAGALGAIASTSAFAVTA 1405

Qy  1431 GALSVTSDITSIVSGALEDASPKASSILGWVSMGMGAAGLAESAI---------KGGTK 1480
         || : :|:| |: ||| . :.||:|:|| |:|  :|||||::|:|         | |:
Db  1406 TVIGLAADSIGIASAALSEKDPKTAGILNWISTGLGVLSFGISAITFTSSLIKSARSGSQ 1465

Qy  1481 LATH--LGAF-AEDGENALLKSTSESSRIKWGVTRS 1513
         :|::   | | :.  |   . ||:||    :  .  :
Db  1466 VASTSVIGSVPIEFGEIA-----SRSSR-RWDIALS 1495
```

RESULT 3

Q883V9

```
ID   Q883V9;            PRELIMINARY;      PRT;  1552 AA.
AC   Q883V9;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   YD repeat protein.
GN   OrderedLocusNames=PSPTO2239;
OS   Pseudomonas syringae (pv. tomato).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC   Pseudomonadaceae; Pseudomonas.
OX   NCBI_TaxID=323;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DC3000;
RX   MEDLINE=22834015; PubMed=12928499; DOI=10.1073/pnas.1731982100;
RA   Buell C.R., Joardar V., Lindeberg M., Selengut J., Paulsen I.T.,
RA   Gwinn M.L., Dodson R.J., DeBoy R.T., Durkin A.S., Kolonay J.F.,
RA   Madupu R., Daugherty S.C., Brinkac L.M., Beanan M.J., Haft D.H.,
RA   Nelson W.C., Davidsen T.M., Zafar N., Zhou L., Liu J., Yuan Q.,
RA   Khouri H.M., Fedorova N.B., Tran B., Russell D., Berry K.J.,
RA   Utterback T.R., Van Aken S.E., Feldblyum T.V., D'Ascenzo M.,
RA   Deng W.-L., Ramos A.R., Alfano J.R., Cartinhour S., Chatterjee A.K.,
RA   Delaney T.P., Lazarowitz S.G., Martin G.B., Schneider D.J., Tang X.,
RA   Bender C.L., White O., Fraser C.M., Collmer A.;
RT   "The complete genome sequence of the Arabidopsis and tomato pathogen
RT   Pseudomonas syringae pv. tomato DC3000.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:10181-10186(2003).
DR   EMBL; AE016863; AAO55755.1; -.
DR   TIGR; PSPTO2239; -.
DR   InterPro; IPR000977; DNA_ligase.
DR   InterPro; IPR006530; YD.
DR   Pfam; PF05593; RHS_repeat; 6.
DR   TIGRFAMs; TIGR01643; YD_repeat_2x; 3.
DR   PROSITE; PS00697; DNA_LIGASE_A1; UNKNOWN_1.
KW   Complete proteome.
SQ   SEQUENCE 1552 AA; 175713 MW; 8DC10DA1BFE37BF1 CRC64;

Query Match    18.2%;  Score 1614;  DB 2;  Length 1562;
Best Local Similarity 30.3%;  Pred. No. 9.3e-77;
Matches  490;  Conservative 239;  Mismatches 700;  Indels 190;  Gaps 51;
```

```
Qy    14 MSDNEFFTQANNFTSAVSGGVDPRTGLYNIQITLGHIVGNGNLGPTLPLTLSYSPLNKT 73
          | .. : :::|:| | :.:|||||||.|.: |.| |:||||:||| .||.|||.||| :
Db     1 MTTSTSVHSNAFNFMSYLQSGVDPRTGQYTVSINLPEVKSNGLRGPVPLVLSYNPLNVQ 60

Qy    74 DIGEGIGFNFGLSVYDRKNSLLSLLSTGENYKVI-STDKTVKLQQKKLDNLRFEKDLKENC 132
          | | |.|: :.|| ||  . .||| |||| |:.  |.. : ..:|||| : .|:|  .
Db    61 DSGFGLGWNLQLSQYDPGTRIVSLGSGETPKVEGSLGDQLWMPEKKLDSFHFYKQ-DDTR 119

Qy   133 YRIIHKSGDIEVLTGPNNNAFDLKVPKKGLLNPAGHAIYIDWNFEATQPRLNRIYDDLDGH 192
          ||:.|||| :| ||. || | |.| .   | :|:|| : :||  | :|:: ::|||||| 
Db   120 YRVVHKSGGQVBELEVLGSLGNRIALPVRIYSPEGHGITLHYASFGAYQMLSEVVDD--DGQ 178

Qy   193 DIPILNLBYQGLIKTILTLFPGQKEGYRTELRFLNRQLNSIHNFSLGNENPLTWSFGYTP 252
          . ::. . .  :||: .| |..|.|  .  :: .: .:.| .| ||.:  . : ||:|
Db   179 VILTITRD----STSVRLLLYGAPKADAEFVMILSGSNRNVARIELPTANKASWRFTYSI 234
```

```
Db  1270 QLGASQNSTMRGDGYLLAEQQGSDALLFSTSISNSVLSEVHPDGVSNRSYTVIGHSSGD 1329
         ::          :   :||:      :    :          ::    :: ||
Qy  1314 ANDASILGYNGERADPVSGVTHLGNGYRSYDPTLMRFHTPDSLSPFGAGGINPYSYCLGD 1373
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1330 DPPAGRLGYNGELHETDTGWQLLGNGYRAYNPVLMRFHSPDSWSPFGEGGLNAYAYGEGD 1389
Qy  1374 PINRSDPSGH----------------LSWQAWTG------------IGMGIAGLLLTIATGG 1407
         ::| :|||                 :| |||
Db  1390 SVNGVDPTGHGWFKLPRIFRTRRLEKQAWKADKKEFWRLIEQDIKNEGLEGRYAQAYRDL 1449
Qy  1408 MAIAAAGG---IAAAIASTSTTALAFGALSVTSDITSIVSGALEDAS--PKASSILGWVS 1462
         ||::|::      :: ::        :|  |  ::    :   :
Db  1450 QAKSAKNNKAYISSLVKHRREKALE--AETLKANTGDVMSSSREESSITPKAEASSGSIS 1507
Qy  1463 MGMGAAGLAESAIKGGTKLATHLGAFAEDGENALLKSTSESSRIKWGVTRSLDREIVRN 1521
         ::          :: ||          :        :
Db  1508 ----------------NLGGFSGSDRDVGLRNVSSA---KSEVLRRQD--FVRN 1540
```

RESULT 4
Q7NWU2
ID   Q7NWU2   PRELIMINARY;   PRT;   1385 AA.
AC   Q7NWU2;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Nematicidal protein.
GN   Name=rhsA; OrderedLocusNames=CV1887;
OS   Chromobacterium violaceum.
OC   Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC   Neisseriaceae; Chromobacterium.
OX   NCBI_TaxID=536;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 12472 / DSM 30191;
RX   MEDLINE=22882880; PubMed=14500782; DOI=10.1073/pnas.1832124100;
RA   Vasconcelos A.T.R., de Almeida D.F., Hungria M., Guimaraes C.T.,
RA   Antonio R.V., Almeida F.C., de Almeida L.G.P., de Almeida R.,
RA   Alves-Gomes J.A., Andrade E.M., Araripe J., de Araujo M.F.F.,
RA   Astolfi-Filho S., Azevedo V., Baptista A.J., Bataus L.A.M.,
RA   Batista J.S., Belo A., van den Berg C., Bogo M., Bonato S.,
RA   Bordignon J., Brigido M.M., Brito C.A., Brocchi M., Burity H.A.,
RA   Camargo A.A., Cardoso D.D.P., Carneiro N.P., Carraro D.M.,
RA   Carvalho C.M.B., Cascardo J.C.M., Cavada B.S., Chueire L.M.O.,
RA   Creczynski-Pasa T.B., Cunha-Junior N.C., Fagundes N., Falcao C.L.,
RA   Fantinatti F., Farias I.P., Felipe M.S.S., Ferrari L.P., Ferro J.A.,
RA   Ferro M.I.T., Franco G.R., Freitas N.S.A., Furlan L.R.,
RA   Gazzinelli R.T., Gomes E.A., Goncalves P.R., Grangeiro T.B.,
RA   Grattapaglia D., Grisard E.C., Hanna E.S., Jardim S.N., Laurino J.,
RA   Leoi L.C.T., Lima L.F.A., Loureiro M.F., Lyra M.C.C.P.,
RA   Madeira H.M.F., Manfio G.P., Maranhao A.Q., Martins W.S.,
RA   di Mauro S.M.Z., de Medeiros S.R.B., Meissner R.V., Moreira M.A.M.,
RA   Nascimento F.F., Nicolas M.F., Oliveira J.G., Oliveira S.C.,
RA   Paixao R.F.C., Parente J.A., Pedrosa F.O., Pena S.D.J., Pereira J.O.,
RA   Pereira M., Pinto L.S.R.C., Pinto L.S., Porto J.I.R., Potrich D.P.,
RA   Ramalho-Neto C.E., Reis A.M.M., Rigo L.U., Rondinelli E.,
RA   Santos B.B.P., Santos S.F.R., Schneider M.P.C., Seuanez H.N.,
RA   Silva A.M.R., da Silva A.L.C., Silva D.W., Silva R., Simoes I.C.,
RA   Simon D., Soares C.M.A., Soares R.B.A., Souza E.M., Souza K.R.L.,
RA   Souza R.C., Steffens M.B.R., Steindel M., Teixeira S.R., Urmenyi T.,
RA   Vettore A., Wassem R., Zaha A., Simpson A.J.G.;
RT   "The complete genome sequence of Chromobacterium violaceum reveals
RT   remarkable and exploitable bacterial adaptability.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:11660-11665(2003).
DR   EMBL; AE016916; AAQ59561.2; -.
DR   InterPro; IPR006530; YD.
DR   Pfam; PF05593; RHS_repeat; 6.
DR   TIGRFAMs; TIGR01643; YD_repeat_2x; 3.
KW   Complete proteome.
SQ   SEQUENCE   1385 AA; 152878 MW; 8AAB59C2C91C2BBE CRC64;

Query Match   18.0%; Score 1598; DB 2; Length 1385;

```
Qy   253 IGKNGILGQWITSMTAPGGLKETVNYSNNNQGHHFPQSANLPVLPYVTLMKQVPGAGQPA 312
         :          :   :||:      ::          :: :|          :|:||
Db   235 IRGHSC----IASVDTPVGGHEDVFYQDS--GHQPPLSAGREPLPRVTRHLTTPGFLQPE 288
Qy   313 IQAEYSY-----TSHNYVGGGSNGIW-NNKLDNLYGLM--TEYNYGSTESRRYKDKBGHD 364
         :            ||:|||    :    :|||||||:|   ||||||:::|||||  | |
Db   289 VDVRYAYKDGAGRERNFLGAGLDIAWEDNGLDNLYRYLGAAPYLYSSTETLRVNDVD--- 345
Qy   365 QIVRIERTYNNYHLLTSECKQQNGYIQTETAYYAIIGHNFDSQPSQFQLPKTKIETWR-- 423
         |||:|||::|:|||||:|  ::| :| | |:|||||||||  |:|||| ||| ||||
Db   346 -VRSIERVFNQFHLLALETTRQNLSILEVDTRYYIEEGKPFDLQPNYCQLPKEVRTTWRL 404
Qy   424 SADNSY-RSEITETTFDESGNPLTKVIKDKKTQKIISPSTHWEYYPPAGEVDNCPPBPYG 482
         |:|:|  | |:|  |:|:|||:|:   ||  :   : ||:||::|||  ||:|||  |:|
Db   405 SPDGSVPRTEIVSSDYDSYGNLLAQTQANGVTETS-------EWYSVSGE-DGCPPDPDG 456
Qy   483 FTRFVKKIIQTPYDSEFKDDPEKFIQYRY----SLIGSQSHVTLKIE-----ERHYSATQ 533
         |.|.:||  :  .:| :   ::||: |||     |:|| |: |:||      ||||||:|
Db   457 FVRTLKAKSVVPAQSDYGHALVLVTRYRYKALPALAGSGQNLWLAAESETLLQQTTDGEK 516
Qy   534 LLNSTLFQYNTDKS----ELGRLLKQTECTKGENGKT----YSVVHKFTYTKQDDTLQQSH 586
         :|:|| :|| |     ::|||:|:|   |  ::||||     ||:: |:|| |||:|||:
Db   517 ELQQTTYTYIEDNPYDAFQYGRIRHQSVTLEGLSTTTDYRYDLLQDPDF--DQTVQQTV 573
Qy   587 SITTHDNFTIHRSQVR---SRYITGRLFSDTDTKDIVTQMSYDKLGRLLTRTLN-SGTPYA 642
         .|:|:|:| |:||:|    |||: |||||:|||||:|||:|||||||||||||  ||||||
Db   574 QIVT--GFDMTQKVIRLEHSLFTGEPLLNRDDNDVEIRYDYDNLRRVVSETVSPNKEBYK 631
Qy   643 NTLTYDYELNNLQDDNRPPFVITTTDVNGNQLRNEFDGAGRHVSQCLKDSDGDGKFYTIH 702
         .|:|::|:| :|:::||  |: |||| |||||||:||||||||:|| ||:||:|| | :|
Db   632 ATRHYEYQLCAVKTDQAEQRLF---DVKNVQTTSRFDGLGRVIYEARADADNPPDVHRRLD 688
Qy   703 TQQ-----YDEQGRHHTSTYSDYLTNGRQQTDPDKVHLSMSKSYDNWGQIANTHWSYGVS 757
         |:|       :|||:||||||||| |||:|||||||:| ||:|||:|||||:||||||||
Db   689 LRQTYEAAYDAWGDKVEETSYDWL-------DQQKRALTNYFEYDDWDQQLSVTGPDGVT 741
Qy   758 EKITVDPITLTAT----KQLQSNSNNVQTGKEVTTY-TPSQQPIQITLFDEAGHLQSCH 811
         :::|.| |::|     ||||:||:|||||:|:|| | ||:|||||||:|||||||||:|
Db   742 TIEQTDPVGTQASNGPIQRRWTESNDGLQTSEVSETWLNLFDEPTRSVRLDRLDWLSEPV 801
Qy   812 TLIR---DGWDRVRKETDAI---GQCTIYQYDNVNRVIQITLPDGTIVNRKYAPFSTDTL 865
         :|:|   ||:|||:|| :|    |:||:|||||::|||:|||||| :|||||||:|:|:|
Db   802 SLSRYQYDGLGRLVKEVSGLPIRERSTTYGYDVFDRVTANTLPDGAVVRRRYAPHSGEDL 861
Qy   866 ITDIRV-NGIS---LGQQTFDGLSRLTQSQDGGRVWAYTYSAGNDQCPSTVITPDGQFIH 921
         : |: | ||:|   ||:| |||||:::|:||||||:|:|||:|:|:|||:::|||||:|
Db   862 PAWIGVDHNGKSSVLGEQKFDGLDRIVTSITTGSREREISYTSDLMQ-PKTVKLPSGRQID 920
Qy   922 YQYQPELDDAVLQ------VASNEITQQFSYNPVTGALLKAVAEGQSLTPIYYPSGRLKM 975
         |:|:|||.| :|       |:||:|||||||||:|| |:|:| ||||||||||||||:|
Db   921 YDYLPELGDEPLKRTQSDTVARLKLTADYTYDPQNARLIGSSBQGEELQREYYSTGSLKS 980
Qy   976 ENINDM---KKMSYLWTLRGLENGYTDLTGTIQKISRDTHGRVTQIKDSSIKTTLNYDDL 1032
         |::|:    |:||:: ||||:|||:||:| :||:|:|   |::   :||:::|:|:|:||
Db   981 EQRTSQGIENTMHYRYSRLGLPLSYIDVLGEQLSVDDFGRLAQTSLGEVVSDFTYDTF 1040
Qy  1033 NRHIGSQVTDLATGHMLTTVEFDGLNREIGRKLCDSSGHTLDIQQSWLKTQQLANRIVK 1092
         .|: :::|.|| :|::::::|:|:|:||:||:|  :::| |:: :|::|::|:|:|::
Db  1041 GRTASIATLDSSSNGQVVISLEYDAQGREAQRTF-TINGANQQMVQVYDDVDQMVKRTLS 1099
Qy  1093 LNGVLQRTEQYSYDSRNRLNQYKCDGAEBCPTDKYGHSIVTQNFTYDIYGNITACHTTFAD 1152
         ::| ::|::::.||:||||.|: |  :  ||: ::|:||::::::||:|||:|: : | |
Db  1100 EGAVIIRREHYGYDLCGRLTQYDCSGKQRPVDFYGMTISRQVFSFDGLNNLLTLVTTTF-D 1158
Qy  1153 GTEDHATFKFANPTDPCQLTEVHHTH-----------PDMPDNIRLKYDKAGRVIN 1197
         |.::::: :|::| | .||| |::::            ::  ::||| |||||:::
Db  1159 GGRNRARY-FYEGIDPAQLTRVTNTQLAWVNARLIPVNKNDSTYPPEIRLYTYDPDGNL-- 1215
Qy  1198 ITDNHGNTENFTYDTLGRL---QNGQGSV-YGYDPLNRLVSQKTDTLDCELYYRETMLVN 1253
         |||:  ||:||:||||||    ::| | | |.||||:||:::::|:| ||||:|  :|
Db  1216 ITDEADRL--LSYDPLGRLLEVSMPAGDVRYRYDPQDRLAGETGE---QRFYRDGVLAS 1269
Qy  1254 EVRNGEMRLLRTGETIIAQQRASKVLLTGTDSQQSVILTSDKQNLSQEAYSAYGKHKST 1313
```

```
Qy    986 PIQQTYSYLLPMGLLHEABEVGGAQSAWD-RWPSGRLR-EETHDIRSGGKKKAHYRSLT 1043
              :                     |||              :|| |::|||
Db    992 GLENGYTDLTGTIQKISRDT---H-GRVTQIKDSSIKTTLNYDDLNRHIGSQVTDLATGH 1047
           ::    :|     :||:  ::|     |      ::|::: :|:|  |::: |   |
Db   1044 GNLEGGADIDGAAHARSYETAAAHVGKLIEIADAAVTVTLAYDGLQRLCSWTARD-GRGH 1102

Qy   1048 MLTTTVEFDLNREIGRKLCDSSGHTLDIQQSWLKTQQLANRIVKLNGVLQRTEQYSYDS 1107
           :|.||:||| ||| :|:|  ::|||| ::.|  |:|     :|||||::||||
Db   1103 ALATTLEFDSLGRETKRTLAAESAEAETLSQEWYPNGQLHQRKRSEGGKPFCDETFVYDA 1162

Qy   1108 RNRLNQYKCDGAECPTDKYGHSIVTQNFTYDIYGNITACHTTFADGTEDHATFKFANPTD 1167
           |||| ||  |||  | |.|||:|  |:| :|:||||  |||||| :||| || ||||  |
Db   1163 RNRLKDYAASGPGLPKDAYGNAIRGQKFEPDAFNNIRKCTTVLDGGSENVGEYLFENPAD 1222

Qy   1168 PCQLTEVHHTHPD--MPDNIRLKYDKAGRVINITDHNGTENFTYDTLGRL---QNGQGS 1222
           ||||| | |  |    |||:||||| ||: ||::|  |:|| ||| ||||        |
Db   1223 PCQLTKVTNSALDKGYPPAIELKYDQAGRLER--DEAG--RRLSYDALGRLARVEGGGGS 1278

Qy   1223 V-YGYDPLNRLVSQKTDT--LDCELYYRETMLVNE--VRNG-------EMRLLRTGETI 1270
             |||||.:||| |       :   |||| :|:||          ||| | |
Db   1279 ASYGYDAHDRLVCQRVETSGMDHRLYYRANRLVNEWMTRSGQAPGADDDRVRLVYAAGSC 1338

Qy   1271 IAQQR----ASKVLLTGTDSQQSVILTSDKQNLSQEAYSAYGKHKST 1313
            |||      ::||| |::|: | |::|:|::| ||:| ||  ::|
Db   1339 AAQVNEGGDGSVAALMGTDGKGSIVSQAEGGQAKHYAYTPYG-HQSS 1384
```

RESULT 5

```
Q7N4A7

ID   Q7N4A7          PRELIMINARY;      PRT;   1775 AA.
AC   Q7N4A7;
DT   01-MAR-2004 (TrEMBLrel. 26, Created)
DT   01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Complete genome; segment 9/17.
GN   OrderedLocusNames=plu2442;
OS   Photorhabdus luminescens (subsp. laumondii).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC   Enterobacteriaceae; Photorhabdus.
OX   NCBI_TaxID=141679;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TT01;
RX   MEDLINE=22957627; PubMed=14528314;
RA   Duchaud E., Rusniok C., Frangeul L., Buchrieser C., Givaudan A.,
RA   Taourit S., Bocs S., Boursaux-Eude C., Chandler M., Charles J.-F.,
RA   Dassa E., Derose R., Derzelle S., Freyssinet G., Gaudriault S.,
RA   Medigue C., Lanois A., Powell K., Siguier P., Vincent R., Wingate V.,
RA   Zouine M., Glaser P., Boemare N., Danchin A., Kunst F.;
RT   "The genome sequence of the entomopathogenic bacterium Photorhabdus
RT   luminescens.";
RL   Nat. Biotechnol. 21:1307-1313(2003).
DR   EMBL; BX571867; CAE14816.1; -.
DR   PhotoList; plu2442; -.
DR   InterPro; IPR006530; YD.
DR   Pfam; PF05593; RHS_repeat; 6.
DR   TIGRFAMs; TIGR01643; YD_repeat_2x; 3.
KW   Complete proteome.
SQ   SEQUENCE 1775 AA; 200136 MW; 0621A9FC06D8801D CRC64;

Query Match             17.8%;  Score 1583;  DB 2;  Length 1775;
Best Local Similarity   28.2%;  Pred. No. 5.1e-75;
Matches  511;  Conservative  280;  Mismatches  727;  Indels  294;  Gaps  65;

Qy     22 TQANNFTSAVSGGVDPRTGLYNIQITLGHIVGNGNLGPTLPLTLSYSPLNKTDIGFGIGF 81
           :||:||:|.:||||||||||:||||: .: | |:|:  ||:|:|: :|||
Db     19 SQATNFTGAMGGGVDPRTGLFTPQIPITQLGGNDLWGPDLAITLITYNPLTYLNTGYGLGF 78

Qy     82 NFGLSVYDRKNSLLSLSTGENYKVIETDKTV------KLQQKKLDNLRFEKDLKENCYRI 135
           .|  . ||: :.:|:|:||| |:|::||:|        |  .|  :  | :|   |  :
Db     79 SDNFTRYDTQTQVLTLATGEIYHRVEKANEVDGQAWTFHHAKPAHFKVKKE--KDAFWV 136

Qy    136 IHKSGDIEVLTGFNNNAFDLKVPKKLLNPAGHAIYIDWNFEATQP----RLNRIYDDLDG 191
```

```
Qy     14 MSDNNEFFTQANNFTSAVSGGVDPRTGLYNIQITLGHIVGNGNLGPTLPLTLSYSPL-N 71
           |:||| . ||| ||:|||.|||||||||: .:  ||| :|:|:| |||:|::||: |
Db      1 MADNNTGIYSQAFNFSSAVNGGVDPRTGLFSFNFDGILSANQGLTPALPLRIAYFPLAAG 60

Qy     72 KTDIGFGIGFNFGLSV-YDRKNSLLSLSTGENYKVIETDKT-VKLQQKLDNLRFEKDLK 129
            :| | |.||:.|  : |||. :||  ::||  | ::::    : :||  :  | :|
Db     61 ESSDGLGKGFSMGGFIKYDRLRRLLILASGERYRLLDEDSAPVFIRQLKLDVVRVHRTVA 120

Qy    130 ENC--YRIIHKSGDIEVLTGFNNNAFDLKVPKKLLNPAGHAIYIDWNFE-ATQPLNRIY 186
               :||::.|.| |  |:  .|||:|:: |  :.||| |:|:::|:|:  .  |:||:
Db    121 DGVRRYRIVILKNGEIHHLSA--PWGPDIYVPEKIQSPLGHALYLSWDSPCGAGRLRLKEVR 178

Qy    187 DDLDGHDIPLLNLEYQGL--IKTILTLFP-GQKEGYRTELRFLNRQLNSIHNFSLGNENP 243
           |:   ||   :  :|       | : |.|  .:   |  :| : |:  :|||:|:|:|
Db    179 DE---EKRTLFRIDYPNADGERVAITQWPDSDDEKVALELYFQNGYLHRIVNKSLSGNGD 235

Qy    244 LTWSFGYTPIGK--NGILGQWITSMTAPGGLKETVNYSNNNQGHHFPQSANLP------V 295
           : |:.|| . .    : | : ::.:||| ||::.|.|          |:|  .
Db    236 VEWTLGYETDSKVADAVGGLLLNELTAPTGLAQRVRYE--------PLCMKLQGERSDFG 287

Qy    296 LPYVTLMKQVPGAGQPAIQAEYSYTSHNYVG-----GGSNGIWNNKLDNLYGLMTEYNYG 350
           ||.|.| . |||||||| : :|.|: .||.|      ||:  |:   ::|:: :  ||
Db    288 LPAVVLHSLVPEAGQPAINTHYEYSPANYLGYGASFKGSQG---GADELFDIQVPYTYQ 343

Qy    351 STESRRYKDKEGHDQIVRIERTYNNYHLLTSECKQQNGYIQTTETAYYAIIGHNFDSQPS 410
           ||| : .|. . .|:| | :|.|||:|||.|| :::|| : |:|||:  ::|::::: ||
Db    344 STEKLLDKSLNPAKPIRTTVRKYNNFHLLVSEEVREGACVFRQETAYPAKVGQSYEAQPA 403

Qy    411 QFQLPKTKTETWRSADNSYRSEITETTFDESGNPLTKVIKDKTTQKIISPSTHWEYYPPA 470
            |||| .|::.|.:| .||| | ::  ||:.|: | |  ::         .:: |||||
Db    404 TFQLPVRQTMSWEAAGRSRRESIS-FEYDDAGQLIKQTMSD-------GSITVLEYYPPA 455

Qy    471 GEVDNCPPEPYGFTRFVK-KIIQTPYDSBFKDPEKFIQYRYSLI----GSQSHVTLKI 524
           || .:||.: .|| |::| || :.  .:.   :: .  | :: |    ||:  .::
Db    456 GEPGHCPADAEGFGRYLKSKTVFPSVSAEYGDEMAMRTEYVFRSIRTPGSAHAEAILQQ 515

Qy    525 EERHYSA-------------------------TQLLNSTLFQY--NTDKSELGRLLKQTECTKG 561
           :: .  .                          |.|    |::    :.:.:.|:|: :::. |
Db    516 TVSHYAGMPGFKARAAMLGKPSWEAYQPRQLAKESYDYLDAPAQKDHGRIKKRIAVVVG 575

Qy    562 ENGKTYSVVHKFTVT------KQDTLQQSHSITTHDN----------FTIHRSQVRSRYTG 607
           |: .  .:| :.     |:|    .: :    :| ::           || |||:|| |:|
Db    576 EDRTPYEMVQDFVFEPVRSGNREVALKQTVSVTVKEDPLEKKDGKLQKVSSTRVLSVLTG 635

Qy    608 RLFSDTDKDIVTQMSYDKLGRLLTRTLNSG-TPYANTLTYDYELNNLQDDNRPPFVITT 666
           ||:| || ::  . .|||.||| |:| :.   .  .  ::.|  ::  :  .    ::
Db    636 RLLSETDVLGNTVAYGYDPLGRLKTQTAHPDLKAYRAIARWDYIWPSTK--NGTPAMAIH 693

Qy    667 TDVNGNQLRNEFDGAGRHVSQCLKDSDGDGKFYTIHTQQYDEQGRHHTSTYSDYLTNGRQ 726
           || .||| |..  | |||: :|  | |.| | . .  .  :|.||: |:  |||:..  |
Db    694 TDALGNQTRTSHDGLGRMIREEACDRDGLGWKTVRTHLYDEAGRQARTTVTDVVHD--R 751

Qy    727 QTDPDKVHLSMSKSYDNWGQIA-NTHWSYGVSEKITVDPITLF-ATKQLQSNSNNVQTGK 784
            .. ::|||:|:|||.:|||::  .| |:|:.:. : .| ::|  | :|.:..:.  :.|
Db    752 EGKPVTLSKTTEREWDSWGQLSVERELETGLASRQEIDPIAQTVATWQAGTDR---CSAK 808

Qy    785 EVTYTTPSQOPIQ---ITLFDEAGHLQS--------CHTLTRDGWD---RVRKETDAIGQC 831
           . ...   .   .    | | :.  : |          |.|| .:|    |:|::|  .
Db    809 YMNFYSKGSHDLERRIVL---AYHLESRSWDAEDKPYSVASWGNDGAHRLRRATDEMSHA 865

Qy    832 TIYQYDNYNRVIQITLPDGTIVNRKYAPFSTDTLITDIRV--NGIS--LGQQTFDGLSRL 887
           | | ||..|| : :.|||:::|.||||||| : |  :.  .  .|     .|:.|.|||||
Db    866 TGYRYDAWGRTVEIVLPDGSAVRKQYAPFSQAALPTQISVADKGVETVAGTQKFPDGLGRL 925

Qy    888 TQSQDGGRVWAYTYSAGNDQCPSTVITPDGQFIHYQYQPELDDAVLQVAS--------N 938
            :.  .||:    .|::    | ::.|.||:.  : |   |   ..  ..        .
Db    926 KRTESGGRPTRFEYASDAASSPRTVTGPDGRVQVAVDDRLGEALKSVAAKAPDHQLGVS 985

Qy    939 EITQQFSYNPVTGALLKA--VAEGQSLTPIYPSGSRLKMENINDM------KKMSYLWTLR 991
```

```
Db  137  LYKTGSREKLTQL-DRANPVAVVSEIYAPSGHKLCVKWNSFVNHNYNYWQLMEVCDAME-  194
         ::::|::|::      ||  ::  : | |    :|: ||      :|||::       ||
Qy  192  HDIPLLNLEYQGLIKTILTLFPGQKEGYRTELRFLNRQLNSIHNFSLGNENPLTWSFGYT  251

Db  195  ---TLLKADLATTEKIEFTWPGSPESYTVTLNMINDLLQTVISAS-----ELTWHLEYE  246
            : ||  ::||  |::         ::|   : ||  |: :|:          |||
Qy  252  PIGKNGILGGWITSMTAPGGLKETVNYSNNQGHPP-----------             288

Db  247  TEGAHKNI--LTKVTTPSGLIEKVVY--HETGHTLPTPKCIQYYPNAWGPGIIRQDPKS  301
           ::  :|   ||||| ||:||: |    |  |::  |::::: | ||: |:||::||
Qy  289  -QSANLPVLPVTLMKQVPGAGQPAIQAEYSYTSHNYVGGGSNGIWN---NKLDNLYGL   343

Db  302  STTATVEHFPVTQHHIIAGSGSPDQVIRVFVFSPENFLGQKNKMDPIPLPQQDNAYLA   361
          :|  ||||:|:  :     :  :| |::    |:|::|||  :|    :  ||:|| |
Qy  344  MTEYNYGSTESRRYXDKEGHDQIVRIERTYNNYHLLTSECK------QQNGYIQYTETAY  397

Db  362  NSEYXYTSTEVREYNEKR-----YCIHREYNKFHLLVSETETVEVTPSRPQKLKETIIKY  416
          :|:| :|||:|  | ::       : |  :|::||| |||||:|||||  ||||||:|||
Qy  398  YAIIGHNFD--SQPSQFQLPKTKIETWRSADNS-----YRSEITETTFPDESGNPLTKVIK  450

Db  417  YADVGKSFDDNKQPPQFLMPNTVETIWHNPESSASTTQRRKETTQWEYNAQCNLISMTLP  476
          ||  |:||::||:| ||  |||:|||| |  || ::|||:|||:|||:| |:|||||||
Qy  451  DKKTQKIISPSTHWEYYPPAGEV--DNCPPEPYGFTRFVKKI-IQTPVD-SEFKDDPEK   505

Db  477  DNTTTKT------TYYAPDGEETTDTHCPAEPNGFERPFIKEIAVEAPSPLTKTKITILR  529
          |:|||:|      |||||||||||||||||||||||| ||||||||||||||||||||||
Qy  506  FIQYRYSLIGSQS----------HVTLKIEERHYS-----ATQL---LNSTLFQYNTD   545

Db  530  KVTYNYKSYDTYSPPKNNQASVVKSMVLPLSETHYSRRDCCADHLEKVKINTVSFYENTQ  589
          |:|||:||:|||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  546  KSELGRLLKQTECTKGENG-KTYSVVHKFTYTKQDDTLQQSHSITTHDNF-----TIHRS  599

Db  590  NAFLNGRVEQRNSVLFENGNQTRSYTEDYSWSNENKNGASCIKCKTTGSGKCCGTPPVSHE  649
          : || || :||:|:|:||||||||||||||||||||||||||||||||||||||||||||
Qy  600  QVRSRYTGRLFSDTDTKDIVTQMSYDKLGRLLTRTLNSGTPYANTLTYDYELNNLQDDNR  659

Db  650  QYWSRSTGRLIFQKDAQDNNTVFQYDTIGHLISSTINADTAYEKTVKIAWSYSNKR----  705
          |:|||||||| |:||| |:|||||||||||||:|:|||||||||||||||:||||:
Qy  660  PPFVITTDVNGNQLRNEFDGAGRHVSQCLKDSGDGKFYTIHTQQYDEQGRHHTSTYSD   719

Db  706  --VTVTQTDIHENRYIITEMDGLGRPLKKSYSPAGYQGKQFDMERYQYNPLGQLCQAISCD  763
             |:|:|||:||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  720  YLTNGRQQTDPDKVHLSMSKSYDNWGQIANTHWSYGVSEKITVDPITLTATKQLQSNSNN  779

Db  764  HML-FENKSEKKCSSMTVSLEYDDWGHNSRLRYSDGTSVKNTFDPIKMTSEHQRVSNDES  822
          :|:  :|:|:||| ||||:|||||||||||||||||||||||||||||||||:|||||||
Qy  780  VQTGKEVTTYTPSQQPIQITLFDEAGHLQSCHTLTRDGWDRVRKETDAIGQCTIYQDNY  839

Db  823  QSSGIIRTTYNQFGQITATERLTTSRIQQGCWHYLRDELGRL-VSINANGNTTLLAYDAF  881
          |:|||||||||:|||||||||||||||||||||||||||||| |:||||||||||||||
Qy  840  NRVIQITLPDGTIVNRKYA-PFSTDTLITDIRVN--GISLGQQTFDGLSRLTQSQDCGRV  896

Db  882  DRVLKQTFADGTTISMAYENGVSVRMSATPLGINQTPVILGTQILDLGRVIDMESGGRK  941
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Qy  897  WAYTYSAGNDQCPSTVITPDGQ-------FIHYQYQPELDDAVLQVAS--NEITQQFSYN  947

Db  942  IKLDYE-GASPVPDTVTYVKGPTSGDKDIVLHYEYEPKLNNAVTKITAGQDDVDQTFKYD 1000
          ||||:| ||||:|||||:||||:|||||||||||||||||||||||||||||||||||||
Qy  948  PVTGALLKAVAE----------GGSLTPIYYPSGRLKME-----NINDMKKMSYLWTLR-G  992

Db 1001  PKTGLLTEAIEEYKNKNDEMISSKLIFDYTLSGQLASEKLVSYNKNNRAYMHFKTCYSFS 1060
          |:||:|||||:|| :|::: |:|||||||||:||:||||| ||:|::|||:||||||||
Qy  993  LENGYTDLTGT--------IQKISRDTHGRVTQIKDSSIKTTLNYDDLNRHIGSQVTDL-- 1043

Db 1061  FANRPTAIKSTINGPRYIINKYSYDNQGRLETVRNNDIEINLKYDKLSRHVCQSTEYLFD 1120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1044  ATGHMLTTVEFDGLNREIGRK-----LCDSSGHT-LDIQQSWLKTQQLANRIVKLNGV 1096

Db 1121  TINNKITTSLTLDDFGRETKRTIRRHNGLKNNKGITDIHISQTFNEQDKITNKEL-LHGT 1179
          |:||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Qy 1097  LQ-RTEQYSYDSRNRLNQYKCDGAECPTDKYGHSIVTQ--NFTYDIYGNITACHTTFADG 1153

Db 1180  KQISKEVYTYHKRGGLETYTL------MELVGERRITQCSDYKYDHLGNITQ-HSITTEG 1232
          :  ::||:||:|||||||||       :|||||||||||||||||||||||| |||||||
Qy 1154  TEDHATFKFANPT--DPCQLTEVHH-THPDMPDNIRLKYDKAGRVINITD-NHGNTENFT 1209

Db 1233  KTITSTYTYGNGNIQDPCQLIDVSTCTGNNSSSSLRFTYNQRGALVCENDNNTKIRTWT 1292
          |:|:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1210  YDTLGRLQNQOGSV---YGYDPLNRLV--SQKTDT-LDCELYYRETMLVN------     1253

Db 1293  YDSLGRIDTTVDALHKVETRYLFDATNRLLIKRSEKNGTPYHHDLSYCCNSLVHDNYFYG 1352
          |:|||| |:|:|||||||||||||||||||||||||||||||||||:|:|||||||||||
Qy 1254  EVRNGEMIR-------LLRTGETIIAQ--QRASKVLLTGTDSQQSVILTSDKQNLSQE 1302

Db 1353  EKRDNAADRKYNKVGGICLGFSQTPCHQTPTSVASRYTETATDGKGSVIATFQGEDVQHI 1412
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1303  AYSAYGKHKSTANDASILG-----------YNGERADPVSGVTHLGNGYRSYDPTLM 1348

Db 1413  AYSPWG---VTTEQAMVTAGQQPHNTAIEEPRFNGEQWDTASASYLLGNGYRAYRPDLM 1469
          |||| |   |||||||||||||||||||||||||||||||||:||||||||||||||||
Qy 1349  RFHTPDSLSPFGAGGINPYSYCLGDPINRSDPSGHLSWQAWTGIGMGIAGLLLTIATGGM 1408

Db 1470  RFTAPDSWSPFGAGGINAYAYCCGGDPVNLNDPSGHISGWGWANIITGGIGLLLAPFTYGG 1529
          ||:||||:||||||||:|:||||||||||||||||||||||||||||||||||||| :||
Qy 1409  AIAAAGGIAAAIASTSTTALAFGALSVTSDITSIVSGALEDASPKASSILGWVSMGMGAA 1468

Db 1530  SLELGLGVNAARGLT--------ALDAASGVTAIASGALENKNPETSRRLGWMSLGLG-- 1579
          ||:|:|||:||:|||        |||||||||||||||||||||||||||||||||||
Qy 1469  GLAEASAIKGGTKLA-------------THLGAF-----AEDGENALLKSTSE---     1502

Db 1580  --LPSMVIGGYSLAQWVNVRLTNSFRTPYHYPTSLGEVNLSRKSSDWVNARRSLNSGENW 1637
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1503  ------SSRIKWG----VTRSLDREIVRNEBGQVIKDHSRGYTNFMGKGEQAI----L 1547

Db 1638  HSEVGLNGRTIWGSDTKIRGLD---IKYPLEQISRRPSNG--DIVLLSGSHGVQNGDNWL 1692
          |:|||||||||||||||||||    |||||||||||||||  |||||||||||||||:||
Qy 1548  VHGDKDGFLYH--------TEGNKHNGKGPYTRHTPEQLVDYLKDNNIVDLTQGGDKPV 1598

Db 1693  INGSRRGSLLHFPFFKSDMTVYGGSWKGRTVHVRN----LATMSEIDFGTLLNNGNSHI 1747
          |||||||||||||||||||||||||||||||||||    |||||||||||||||||||
Qy 1599  HLLSCYGKSSGA 1610

Db 1748  ILGYCYGRNDQA 1759
```

RESULT 6

```
Q9S6J1
ID   Q9S6J1      PRELIMINARY;      PRT;   773 AA.
AC   Q9S6J1;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
OS   Coxiella burnetii.
OG   Plasmid QpDV.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
OC   Coxiellaceae; Coxiella.
OX   NCBI_TaxID=777;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=R1140;
RA   Radomski K.U., Willems H., Lautenschlaeger S., Jaeger C., Baljer G.;
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF131076; AAD33495.1; -.
DR   InterPro; IPR006530; YD.
DR   Pfam; PF05593; RHS_repeat; 6.
DR   TIGRFAMs; TIGR01643; YD_repeat_2x; 2.
KW   Hypothetical protein; Plasmid.
SQ   SEQUENCE   773 AA;  83759 MW;  4C5FCB04B1CCAC44 CRC64;
```

```
Query Match             17.5%;  Score 1553;  DB 2;  Length 773;
Best Local Similarity   44.6%;  Pred. No. 5.5e-74;
Matches  342;  Conservative 119;  Mismatches 275;  Indels 30;  Gaps 17;

Qy  741  YDNWGQIANTHWSYGVSEKITVDPITLTATKQLQSNSNNVQTGKEVTTYTPSQQPIQITL 800
```

```
DR   TIGRFAMs; TIGR01643; YD_repeat_2x; 2.
KW   Hypothetical protein, Plasmid.
SQ   SEQUENCE   774 AA;  83774 MW;  918A5433D7991BD8 CRC64;

Query Match             17.2%;
Best Local Similarity   44.5%;                    Score 1530.5;  DB 2;  Length 774;
Matches  341;  Conservative  120;  Mismatches  275;  Indels  31;  Gaps  18;
                                                  Pred. No. 8.6e-73;
```

```
Qy    741  YDNWGQIANTHWSYGVSEKIYVDPIITLTATKQLQSNNNVQTGKEVTYTPSQQPIQTTL   800
           |  |||     ::::|:||||| |||   :  | |:   |::   |   ::    |||
Db      4  YDSWGQNHLTVFSDGYQERSVYDPITRRAT--LQPESGSQKLGQQLTEYNLAGLPIKVTQ   61

Qy    801  FDEA-GHLQSCHTLTRDGWDRVRKETDAIGQCTIYQDNYNRVIQITLPDGTIVNRKYAP   859
           |::| |    ||  ||||    ||  |   ||||:||:|| |   ||||| |||  |||
Db     62  YDNSQGTEQGSAHYEYDGLGQLRKETDELGQVTLYEYDHFGRVTQTTLPENTIIQKSYAP   121

Qy    860  FSTDTLITDIRVNGISLGQQTFDGLSRLTQSQDGGRVWAYTYSAGNDQCPSTVITPDGQF   919
           |    ||  |||||| |  |  ||||   | | |   |||  :|   |   |:    | |
Db    122  HSTASLITGISVNNFSMGNQTFDSLERLTETTSGGRTSAFSYENAS-SVPAAVTAPTGET   180

Qy    920  IHYQYQPELDDAVLQVASNEITQQFSYNPVTGALLKAV-AEGQSLTPIYYPSGRLKME-N   977
           :|:|    ::|||     ::   | ||| |||   | |    ||:     | | |:||
Db    181  VSYEYLKELGNAVKKISAPEILKTWDYDALTGAMTSATQAGMIRQMTYYPSGLLKNETS   240

Qy    978  IND--MKKMSYLWTLRGLENGYTDLTGTIQKISRDTHGRVTQIKDSSIKTTLNYDDLNR  1034
           :    |    ||  ||  ||||  ::||   |  ||| |  |     | |:    ||||
Db    241  MPDGAAQKSTAYTYSLAGAPQSYTDVFGVTDVGYTDLQIEDNDIKVSLGYDAFGR   300

Qy   1035  HIGSQVTDLATGHMLTTTVEPDGLNREIGRKLCDSSGHTLDIQQSWLKTQQLANRIVKLN  1094
           :|:|    |   |:|::||  |||||||  : ::|  | |||||| ||||  |::||
Db    301  FTKQQATDKKTGAVLSTTLYDDLNREIKREISASGQSVLVIEQTYQRNHLLKERITQRG   360

Qy   1095  GVLQRTEQYSYDSRNRLNQYKCDGAECPTDKYGHSIVTQNFTYDIYGNITACHTTFADGT  1154
           : |:   : ||||||||   :| :|  | |||| :|  :|:||| |||||  ::  ||
Db    361  RTTLRKEVFAYDSRNRLIEYTCNGEARPQDPYGKAIHRQTFSYDALGNMTKTQTDFSGG-   419

Qy   1155  EDHATFKFANPTDPCQLTEVHHTHPDMPDNIRLKYDKAGRVINITDNHGNTENFTYDTLG  1214
           :||:|::  :|:|| ||  | ::|  |||  :||||||:|:|   | ||| ::| :||||
Db    420  RNTATYIY-SAIDPTQLLKVNNDHSDYPKEITLEYDKAGRMIR--DEAGRT--LRYDALG   474

Qy   1215  RLQ--NG---QGSVVGYDPLNRLVSQKT-DTLDCELYYRETMLVNEVR--NGEMIRLLRT  1266
           |||  ||   :||  |||:||:|||:   | : | :|   |:|||  | :| :|| |
Db    475  RLQQVNGAGAKGGQYAYDALNTLVSQVVQDEPIYDLYRADDLVGEARRDGSSQTRVKS   534

Qy   1267  GETIIAQ---QRASKV-LLTGTDSQQSVILTSDKQNLSQEA-YSAYGKHKSTANDASILG  1321
           |  :|::   |:||:| |||| :|| ||:|||:|:|:|:|  |:||| : :||:|||:||
Db    535  NGCCVGQCTKQGSSNTSRLTTTNQQGSVLSVSEGNHAPQDCIYTRYGYRTPQTETPSVLG   594

Qy   1322  YNGERADPVSGVTHLGNGYRSYDPTLMRFHTPDSLSPFGAGGINPYSYCLGDPINRSDPS  1381
           :||||:||||| ::||||||:|:| :|||: ||| ||||||||||:|| ||||||: ||
Db    595  FNGERLDPVSGTYHLGNGYRAYNPIIMRFNCPDSWSPFGAGGINPYAYCDGDPINRVDPN   654

Qy   1382  GHLSWQAWTGIGMGIAGLLLTIATGMAIAAAGGIAAAIASTSTTALAFGALSVTSDITS  1441
           ||||:|| ||||:| ||:| : |:|:||||||:|::|: |:||:|:  |: | |::  |
Db    655  GHLSSQAELGIGLGVVGLVLAVFTAGTSIAAAGAISAIESASAISLVVGTIGVAADVAS   714

Qy   1442  IVSGALEDASPKASSILGWVSMGMGAAGLAE--SAIKGGTKLATHL  1485
           | |||||| :| || :|||:|:|:| |   :  ::| ||:||  |:
Db    715  IASGALEDANPQASATLGWISLGLGPGAVSGLATAARAGKKLISGL   761
```

RESULT 7

```
O52880
ID   O52880;           PRELIMINARY;      PRT;   774 AA.
AC   O52880;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein orf 774.
GN   Name=orf 774;
OS   Coxiella burnetii.
OG   Plasmid QpRS.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
OC   Coxiellaceae; Coxiella.
OX   NCBI_TaxID=777;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lautenschlaeger S., Jaeger C., Willems H., Baljer G.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Y15898; CAA75841.1; -.
DR   InterPro; IPR006530; YD.
DR   Pfam; PF05593; RHS_repeat; 6.
```

RESULT 8

```
Q88LP7
ID   Q88LP7;           PRELIMINARY;      PRT;   1632 AA.
AC   Q88LP7;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
GN   OrderedLocusNames=PP1882;
OS   Pseudomonas putida (strain KT2440).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC   Pseudomonadaceae; Pseudomonas.
```

```
OX   NCBI_TaxID=160488;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22423060; PubMed=12534463;
RA   Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,
RA   Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,
RA   Brinkac L.M., Beanan M.J., DeBoy R.T., Daugherty S.C., Kolonay J.F.,
RA   Madupu R., Nelson W.C., White O., Peterson J.D., Khouri H.M.,
RA   Hance I., Chris Lee P., Holtzapple E.K., Scanlan D., Tran K.,
RA   Moazzez A., Utterback T.R., Rizzo M., Lee K., Kosack D., Moestl D.,
RA   Wedler H., Lauber J., Stjepandic D., Hoheisel J., Straetz M., Heim S.,
RA   Kiewitz C., Eisen J.A., Timmis K.N., Duesterhoeft A., Tuemmler B.,
RA   Fraser C.M.;
RT   "Complete genome sequence and comparative analysis of the
RT   metabolically versatile Pseudomonas putida KT2440.";
RL   Environ. Microbiol. 4:799-808 (2002).
DR   EMBL; AE016780; AAN67501.1; -.
DR   TIGR; PP1882; -.
DR   InterPro; IPR006530; YD.
DR   Pfam; PF05593; RHS_repeat; 7.
DR   TIGRFAMs; TIGR01643; YD_repeat_2x; 4.
KW   Complete proteome; Hypothetical protein.
SQ   SEQUENCE   1632 AA;  182290 MW;  F07C2A5820160401 CRC64;

Query Match             15.9%;  Score 1416;  DB 2;  Length 1632;
Best Local Similarity   28.4%;  Pred. No. 3.4e-66;
Matches  472;  Conservative 232;  Mismatches 734;  Indels 226;  Gaps 52;

Qy       14 MSDNNEFFTQANNFTSAVSGGVDPRTGLYNIQITLGHIVGNGNLGPTLPLTLSYSPLNKT 73
               :  ::  :    ::   ||||||||:|||:|||::||||||:|  ::||:::|::|||
Db        1 MSTESSLHSNALNFMSFMESGVDSRTGQYTLSIKLPELGANYFSGPELELSLNFNPLNMI 60

Qy       74 DIGFGIGFNFGLSVYDRKNSLLSLSTGENYKVI-ETDKTVKLQQKKLDNLRFEKDLK--- 129
             ::|::  :::   ::: :||||| ||:  |    :| : :  :  |  :|  :: |
Db       61 DSGWGRGWNLSLTQFAPHTQVITTYSGESFKVTGSSGSRLQMQEQRLAHFHFYREPAGPG 120

Qy      130 -ENCYRIIHKSGDIEVLTGFNNNAPDLKVPKKLLNPAGHAIYIDWN-FEATQPRLNRIYD 187
               |||: ||:   :|||:|: :|  |||| || || ||  :||   |  :||| |:||
Db      121 GNARYRVVHRSGLVBILEEMGSANGRIALPVEIYAATGHRIDLQYQPFNSSYMMLSSISD 180

Qy      188 ------DLDGHDIPLLNLEYQGLIKTILTLFPGQKEGYRTELRFLNRQLNS--IHNF 236
                   :||: |||:|:||||:|: |:: ||||  :  | :: |  ::|:  |::|
Db      181 EREEILLEIERDDSRIELRERPYQG---------DNGQPVALYAMN-LINTDWVTSI 227

Qy      237 SLGNENPLTWSFGYTPIGKNGILGQWITSMTAPGGLKETVNYSNNNQGHHFPQSANLPVL 296
             ||:|| ||||  | || :|:|| :|| :||  |||   :|   :|  | |||:| : ||
Db      228 VLPTHELASWRLEYALV--NGLL--CVSKVEPPTGAREYLYY--QDRGHLFPGDAR-PPL 280

Qy      297 PYVTLMKQVPGAQOPAIQAEYSYTS-HNYVGGGSNGIW-NNKLDNLY-GLMTEYNYGSTE 353
             | :|:  |  | | |:: |   |:  :|| |||:|| | :||||||: |:: || :||:|
Db      281 PRVTRHVIEPRGGQPAFQRTYTYTPGLNNFLGYGAGIGWSDNGLIDNLYESKRYDYEYQTVE 340

Qy      354 SRRYKDKEGHDQIVRIERTYNNYHLLTSECKQQNGYIQTETAYYAIIGHNFDSQPSQFQ 413
              :||:|  | |:|  ||:| ::| |:|:: | :||||:|: ||:| :|||||:||||:|
Db      341 TLR--DEDG-TALRDITRTFNRFHLLTSTRTVQNNCVHEV-TMQYNIKDVPFNQQVSTLQ 396

Qy      414 LPKTKTETWRSADNSYRS--EITETTFDESGNPLTKVIKDKKTQKIISPSTHWEYYPPAG 471
             ||:|||| ::|::   :   :|||: |:| ||: :|:| |:: |::||  | ||::   |
Db      397 MPIREQTRWSLADNITRSRLETVETTYDTSGNIIWRKLANGVTET------QEWYGTQA 449

Qy      472 EVDNCPPEPYGFTRFVKKIIQTPYDSEFKDDPEKFIQYRYSLIG--SQSHVTLK----- 523
             |    ||: ||  | :||::||:|:  :::   :| |  :: ||  :||| ||
Db      450 E-DDYPGDANGFVRHLKSKTTTPASSGRGQAPVLTQHYRYKALAPLAGNAITLNPMIVEH 508

Qy      524 ---IEERHYSATQLLNSTLFQYNTDKSBELGRLLKQTECTKGENGKTYSVV---HKFTYTK 577
                ||:| |::|  :| || |::|| | ||||:|| | |:| || ||:|   |||::|
Db      509 SETLTHAHDPAHPLEEKIYYYLDAPKSSL------RHGRRYQEVVKRNKLETTT 556

Qy      578 Q---------DDTLQQSHSITTHDNFTIHRSQVRSRYTGRLFSDTDTKDIVTQMSYD 625
                       |:|::|::|  |::|||::::||:: || :    |:|::|||| ||
Db      557 QYQFNSLLDPLGGHQVLETRKTLFGYDGAQRSTVQRRSLLHGEKLYELNENGVHTQWAYD 616

Qy      626 KLGRLLTRTLNSGTPYANTLTYDYELNNLQDDNRPPFVITTTDVNGNQLRNEFDGAGRHV 685
```

(Right column, continuation:)

```
Db      617 ALRRVTEERVSPETPFEAKKRYDYQLCASDADIARARV----TNARGITTETELDGLGRPT 673
             ||:  ::   :                              ::      |||  ||   ||
Qy      686 SQCLKD--SDGDGKFYTIHTQQYDEQGRHHTSTYSDYLINGRQQTDPDKVHLSMSKSYDN 743

Db      674 RESRDNVLEARPGAFYBILAIQYDAYGNRIQESVTDWLQ--AQQ--YLHLVTKTRYDD 727
             :|:  ::   |::                               :   |||::||:|:
Qy      744 WGQIANTHWSYGVSEKITVDPI---------------------TLTATKQLQS-NSNNVQ--- 781

Db      728 WGEQCCTIGPDDVEQHNVLDPIGNADHQGAIRYSWREGGRLKSRPQVQNRNGMVRSTRQ 787
             ||| :     :  |  ::| :  || |  ||  :   | :  : | |:|  ||::|:|
Qy      782 ---TGKEVTYTPSQQPIQITLFDEAGHLQSCHTLTRDGWDRVRKETAIGQCTIYQYD 837

Db      788 ARAISGKTETWLNLFNKPVLRKRLDALGELGIGERRYSYDGLGRTLTETDERNHTTAFSYD 847
             || :| ||| :|::|| |::|::|| |:::  |||||:||| ||: |||:|| |||: ||
Qy      838 NYNRVIQITLPDGTIVNRKYAPFSTDTLIT------DIRVNGISLGQQTFDGLSRLTQS 890

Db      848 AWGRMLTTBQPNKTLLTRTYAEHSAEDLPTRLVVTPANVQLPARQIGAQYFDGLDRLIGT 907
             ||  |:| ||||:||:||: |:|:|| | | |  ||||:| ||:|| |:||| ||| ||
Qy      891 QDGGRVWAYTYSAGNDQCPSTVITPDGQFIHYQYQPELDDAVLQVASNEITQQPSYNPVT 950

Db      908 TTGDRTEYFLFRDG-ESIPRQRINPAGETIEQDYNLQLTNEPISNTAPEHASFAYDPVS 966
             ||| | |:| :|:  |:|  :||  |:|| : ::||:|::|  :: |:|: :  |||| |
Qy      951 GALLKAVAEGQSLTPIYPYPSGRLKMENINDMKKMSYLWTLRGLENGYTDLTGTIQKISRD 1010

Db      967 ARLLSDDNRQCARRFEYNKANQLTAEHWED-KRDGKTWSRIHSSTLQDRLKNTHEPYGED 1025
              |::|:|:|| || |:|:|| ::| |  |  :| |||| |::|| :|: :|| || |:|
Qy     1011 T------HGRVTQIKDSSIKTTLNYDDLNRHIGSQVTDLATGHMLTTTVEFDGLNREIGR 1064

Db     1026 TTHEYDAQGRLVSTLQGQLQAEFKYDDLGRIELITSHDRASAQALETKIEYDDQDREVKR 1085
             ||: |  :|  :|  || |: |  :|  |: || ||||:|:  |: ::  : |::| |:|
Qy     1065 KLCDSSGHTLDIQQSWLKTQQLANRIVKLNGVLQRTEQYSYDSRNRILNQYKCDGAEFCPTD 1124

Db     1086 TWRQPGQPERTQETVWDKDDLLLSRTLQVGGVSRLVEKFGYDSHARLMYNCTGPDQPRD 1145
             || : |  :||  | |::|: | |::| | :: | | :|| | :|| | |   :  | |
Qy     1125 KYGHSIVTQNFTYDIYGNITACHTTFADG-TEDHATFKFANPTDPCQLTEVHHTHPDMPD 1183

Db     1146 ALGRSIAMQVFNFDAYNNIELTVTSFTGGPAAERATFIHAE-RDPCQLQRIEYTPPRTAP 1204
             ||| |:  |||: |  ::|| |   | ||       |  |  ||||| ||:|  ||  :|
Qy     1184 NIRLKYDKAGRVINITDNHGNTENFTYDTLGRL----QNGQGSVYGYDPLNRLVSQKTDT 1239

Db     1205 NPEFSYDANG---NLTRDE-QARPIRYDSQNRLLGLNDSGAPDTYGYDAGGLLVSRPBAG 1260
             |||||:||    :||||  :||| |||:|:||:|| ||:|| || ||| ||:|||   |
Qy     1240 LDCELYYRETMLVNEVRNGEMRLLRTGETIIAQQR----ASKVLLTGTDSQQSVILTSD 1295

Db     1261 ERTILLHDGQRLRMAVRDSLQTLYLHHDEQALGQQQKTGAQSPLLLHTSASHSVIABSQ 1320
             ||::|| :|:|  |  | |: |: |: ::| | |:: || | | |||:|:| ||: |
Qy     1296 KQNLSQEAYSAYGKHKSTANDASILGNGERADPVSGVTHLGNGYRSYDPTLMRFHTPDS 1355

Db     1321 AGSTRAVRYTAYGERHADDPLLGTLGYNGEALDPDSGWYLLGSGYRAYNPVLMRFHSPDA 1380
             ||| : ||| |  :|  ||  || ||| || |||| | || || || | |||| |::| |
Qy     1356 LSPFGAGGINPYSYCLGDPINRSDPSGHLS-------------------WQ-------AWT 1390

Db     1381 LSPFGAGGLNYYGYCQGNPITFRDPTGHYSIGYSGQSRSLADLNSYSIWRGKALGAIGWT 1440
             |||||||| | |||| |:| | ||:| | |  |: |:|:  |:|||:|||| ||| ||||
Qy     1391 GIGMGI------AGLLLTIATGMA----------------IAAAGGIA----AAIASTS--- 1424

Db     1441 GIGLGILFAAVASVAAVVTGGVAAPAIAAAWAAAGGGISGAGAALSAGVGAIASVSLAV 1500
             |||:|| :||||||: | |||  ||:  |::|  ||:| | || ||||||::||::|| |
Qy     1425 ----TTALAFGALSVTSDITSIVSGALEDASPKASSLLGWVSMGMGAAGLAESAIKGGTK 1480

Db     1501 GIKLAAAVATATLSVTG--TALQTEAILSGHEKRNNIGTILNYSAAILGLAVGAMQLIAK 1558
             ||::||| || ||||  |   || |   : | ::|:|||||: :|| |: ||| |:||:|
Qy     1481 ---LATHLGAFAED---GENALLKSTSESSRIKWGVTRSLDREIV 1519

Db     1559 IPNLAWRAGSYTVTTITDDPLINSGKTLSELK------MARQII 1596
```

```
RESULT 9
Q45948

ID   Q45948      PRELIMINARY;      PRT;      709 AA.
AC   Q45948;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
```

```
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Orf 709.
OS   Coxiella burnetii.
OG   Plasmid QpH1.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
OC   Coxiellaceae; Coxiella.
OX   NCBI_TaxID=777;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Thiele D., Willems H., Haas M., Krauss H.;
RT   "Analysis of the entire nucleotide sequence of the cryptic plasmid
RT   QpH1 from Coxiella burnetii.";
RL   Bur. J. Epidemiol. 10:413-420(1995).
DR   EMBL; X75356; CAA53129.1; -.
DR   PIR; S38241; S38241.
DR   InterPro; IPR006530; YD.
DR   Pfam; PF05593; RHS_repeat; 8.
DR   TIGRFAMs; TIGR01643; YD_repeat_2x; 4.
KW   Plasmid.
SQ   SEQUENCE  709 AA;  79416 MW;  A377B9A2A72CCF86 CRC64;

Query Match             13.7%;  Score 1214;  DB 2;  Length 709;
Best Local Similarity   40.8%;  Pred. No. 5.1e-56;
Matches  287;  Conservative 101;  Mismatches 280;  Indels 36;  Gaps 16;

Qy    554 KQTECTKGENGKTYSVHKPTYTKQDDTLQQSHSITTHDNPTIHRSQVRSRYTGRLFSDT 613
          : :|  ||     |   ||     |   ||  |:   ||  |:  |||| :: :||||||||
Db      3 RRAEVLTSKEGKKYQNQTFAISSQAEHLLQKIDFTGGDGIKTSISREQSRYSGRLLSST 62

Qy    614 DTKDIVTQMSYDKLGRLLIRTLN-SGTPYANTLITYDYELNNLQDDNRPPVITTDVNGN 672
          ||| : |:|   |:   |:|| |  :||||:| |||  ||: | || : |||| ||||
Db     63 DELGNVTQYEYDELGRLLTQTVNASSTTYASTTTRYSYSLETDARGKVTAIKTTVTDPKGN 122

Qy    673 QLRNEPDGAGRHVSQCLKDSDG------DGKFYTIHTQQYDEQGRHHTSTYSDYLTNGRQQ 727
          ||| | ||  |   |   |      ||||  | |    || | || |  |  ||||  :
Db    123 QLRTYDYDGLGRNLKQERLDKDAAVSQKTGTWYITHQQYDTDALGRESKTIITQDVL----R 177

Qy    728 TDPDKVH---LSMSK--SYDNWGQIANTHWSYGVSEKITVDPITLTATKQLQSNSNNVQ 781
          |      |      :: ||| ||  | |  :||| | ||   ||| :|||    :| :| |
Db    178 LDSEVGHAGSVLSTSKVMHYDSWGQNHLTVFSDGYQERSVYDPITRRAT--LQPESGSGQK 235

Qy    782 TGKEVTYTYPSQQPIQITLFDEAGHLQSCHTLTRDGWDRVRKETDAIGQCTTYQYDNYNR 841
          ||  | |||  : ||||| || : |:|   | || ||  : |:||| ||:|||:|| |||
Db    236 LGGQLTEYNLAGLPIKVTQYDSQGTEQGSAHYEYDGLGQLRKETDELGQITLYEYDHFGR 295

Qy    842 VIQTLPDGTIVNRKYAPFSTDTLITDIRVNGISLGQQTFDGLSRLTQSODGGRVWAYTY 901
          |   |||| || |   ||| |  |   :|| :| |||  ||| | ||   |  : ||
Db    296 VTQTTLPENTIIQKSYAPHSTASLITGISVNNFSMGNQTFDSLERLITETTSGGRTSAFSY 355

Qy    902 SAGNDQCPSTVITPDGQFIHYQYQPELDDDAVLQVASNEITQQFSVNPVTGALLKAV-AEG 960
          |   :|   | ||||| : ||   | ||:| |   :|:||| ||| |  |  |   |||
Db    356 ENAS-SVPAAVTAPTGETVSYEYKLEGNAVKKISAPEILQTWDYDALTGAMTSATQAAG 414

Qy    961 QSLTPIYYPSGRLKME-NIND--MKKMSYLWTLRGLENGYTDLTGTIQKISRDTHGRVT 1016
          |:   |||  |    || |  ||  |  :| | |  || |||  |:|:| || || |  ||
Db    415 MIRQMTYYPSGLLKNETSMPDGAAQKSTAYTYSLAGAPQSYTDVFGVTGVTQRYDYDEHGRRI 474

Qy   1017 QIKDSSIKTLNYDDLNRHIGSQVTDLATGHMLITTVEFDGLNRREIGRKLCDSSGHTLDI 1076
          |:  | :| ||:|| || ||:| |||   :  |  ||||||| ||| | ||| | |  ::
Db    475 GIEDNDIKVSLDYDAFGRFTKQQATDKKTGAVLSTTLTYDDLNREIKREISASGQSVLVI 534

Qy   1077 QQSWLRTKQQLANRIVKLNGVLQRIBOYSTSRNRLNQYKCDGAECPTDKYGHSIVTQNFT 1136
          ::  ::|| | |  ||  | | || ||| | :| |  |  | |  ||   |  : |  |
Db    535 EQTYQRNHLLKERITQRGRTTLRKEMFAYDSKRM | |EYTCNGEARPQDPYGKAIHRQTFS 594

Qy   1137 YDIYGNITACHTTFADGTEDHATFKFANPTDPCQLTEVHHTHPDMPDNIRLKYDKAGRVI 1196
          |    |:|  | ||   :   ||  |   |  | ::| |  :|  |  : |:||  |||
Db    595 YDALGNMTKTQTDFSGG-RNTATYIY-SAIDPTQLLKVNNDHSDYPKEITLEYDKAGRMI 652

Qy   1197 NITDNHGNTENFTYTDTLGRLQ--NG---QSVVGYDPLNRLVSQ 1235
          | | : || |  ||||  |:|    ||    ::| |  :||| :|
Db    653 R--DEAGRT--LRYDALGRLQQQVNGAGAKGGQYAYDVLNTLVSQ 692
```

```
RESULT 10
Q88LP2
ID   Q88LP2                  PRELIMINARY;            PRT;   1290 AA.
AC   Q88LP2;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE   Hypothetical protein.
GN   OrderedLocusNames=PP1887;
OS   Pseudomonas putida (strain KT2440).
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC   Pseudomonadaceae; Pseudomonas.
OX   NCBI_TaxID=160488;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22430060; PubMed=12534463;
RA   Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H.,
RA   Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,
RA   Brinkac L.M., Beanan M.J., DeBoy R.T., Daugherty S.C., Kolonay J.F.,
RA   Madupu R., Nelson W.C., White O., Peterson J.D., Khouri H.M.,
RA   Hance I., Chris Lee P., Holtzapple E.K., Scanlan D., Tran K.,
RA   Moazzez A., Utterback T.R., Rizzo M., Lee K., Kosack D., Moestl D.,
RA   Wedler H., Lauber J., Stjepandic D., Hoheisel J., Straetz M., Heim S.,
RA   Kiewitz C., Eisen J.A., Timmis K.N., Duesterhoeft A., Tuemmler B.,
RA   Fraser C.M.;
RT   "Complete genome sequence and comparative analysis of the
RT   metabolically versatile Pseudomonas putida KT2440.";
RL   Environ. Microbiol. 4:799-808(2002).
DR   EMBL; AE016781; AAN67506.1; -.
DR   TIGR; PP1887; -.
DR   InterPro; IPR006530; YD.
DR   Pfam; PF05593; RHS_repeat; 5.
DR   TIGRFAMs; TIGR01643; YD_repeat_2x; 4.
KW   Complete proteome; Hypothetical protein.
SQ   SEQUENCE  1290 AA;  146578 MW;  7081A3F10B287F92 CRC64;

Query Match             11.1%;  Score 1037;  DB 2;  Length 1290;
Best Local Similarity   27.7%;  Pred. No. 3.4e-46;
Matches  375;  Conservative 196;  Mismatches 578;  Indels 204;  Gaps 52;

Qy     14 MSDNNEFFTQANNFTSAVSGGVDPRTGLYNIQITLGHIVGNGNLGPTLPLTLSYSPLNKT 73
           ||   :   :|||| :||||| ||| ||| || | | |  ||| ||| |:||||| |||
Db      2 MSTSTSVHSNAFNFSSYIESGVDPRTGQYTVSIRLPELQGNDLQGPGFELALFYSPLNGE 61

Qy     74 DIGFGIGFNFGLSVYDRKNSLLSLSSTGENYKVI-ETDKTVKLQQKKLJDNLRFEKDLKEN- 131
          |  ||:|  :| | |:|  | |:  | | |||   ||| |  ||: | | :|| ||
Db     62 DSGFGKGWN--LQLTQVRKHIVTLSSGETYKITGKSSVTGRLEMKEQKLRQF--DLYEDP 117

Qy    132 ------CYRIIHKSGDIEVLTGFNNNAFDLKVPKKLLNPAGHAIYIDW-NFEATQPRLN 183
                 | || | : ||:|||| || | |||| ||  :|||| |  | ||| |||
Db    118 PGPGGAARFRVEHRSGLVEVLEVMGSGEGEVALPVELHSPLGHRLHLSYLPFGQGHMRLS 177

Qy    184 RIYDLDGHDIPLLNLEYQGLIKTLILTLFPGQKEGYRTELRP---LNRQLNSIHNFSLGN 240
          ||  | |  |     ||  | |  | |||| | ||||| |||   |:|| :  :||  |
Db    178 EVQDESD----VLLRLQRSDNSRVELLCYPSGGDGGPLARYAMTLSESDNRVSEIILPT 233

Qy    241 ENPLTWSFGYTPIGKNGILGQ-WITSMTAPDGLKETVNYSNNNQGHHFPQSANLPV-LPY 298
          ||| |   |  : | ||| | :  | |   || || ||   |:| |   |   ||||| |||
Db    234 ANQARWRFTY-----EDVLGYLCVRECTTPYGGVERVFY--QDAGHKFPSSAARDKNLPR 286

Qy    299 VTLMKQVPGAGOPAIQAEYST-SHNYVGGSNGIWNNK-LDNLYGLMTEYNYGSTESRR 356
          || |:| |  :   | |  |  | | |  |  || |  |  | |  |   ||  ||||| |
Db    287 VTRHEIDPRFGQAAKVRYEYPGTHNFIGGSSISWSDDGLDNLYKVPBDVTYKSIQI-- 344

Qy    357 YKDKEGHDQIVR-IERTYNNYHLLTSECKQQNGYIQTETAYYAIIGHNFDSQPSQFQLP 415
          | | | ||::|:    |   |  |   | || |     |   :   | ||:| :||| |
Db    345 ---QEVRRQSVRTITRFNRFHLLTEQATARQDKLLQAFTRYADNAG-NFRSQEPYFQLP 400

Qy    416 KTKTETWRSADN--SYRSEITETTFDESGNPLTKVIKDKKTQKIISPSTHWEYYPPAGEV 473
          || ||||  |:   :|| | |||| |||||| || || | |: || ||    ||| ||
Db    401 HDETQRWSLISNPSRQREEKRVTRYDSHGNVLTRLLPNQVLETNV------WYSAEEEG 453
```

```
Query Match             10.5%;
Best Local Similarity   29.7%;       Score 932.5;  DB 2;  Length 982;
Matches  285;  Conservative 142;  Mismatches 357;  Indels 175;  Gaps 33;    Pred. No. 8e-41;

Qy    642 ANTLTYDYELNNLQDDNRPPFVITTTDVNGNQLRNEFDGAGRHVSQCLKDSDGDGKFYTI 701
          ::||:|| ||              ||:|:|||: ::|:||         ||   :  :::
Db     31 SSTLGYRYD------DWNQR--CCTTTDDN-VQTYEYSDPIGSDVHK--------GPIQKT 74

Qy    702 HTQQYDEQGRHHTSTYSDYLTNGRQQTDPDKVHLSMSKSYDNWGQIANTHWSYGVSEKIT 761
           :::|  || :   :|                           |:|:   ||     ::
Db     75 WKQSGDPEGR-------------ISGRSET----------WLNL------FGKPDRIR 103

Qy    762 VDPITLTATKQLQSNSNNVQTGKEVTTYTPSQQPIQITLFDEAGHLQSCHTLTRDGWDRV 821
              | |||||:  ::                   ||:||:|  |    : |||:||||:|
Db    104 ----TLTAGKTGRSRTHSMSRSRNLTT---TEQEL------SRQTFLYDGLGRC 144

Qy    822 RKETDAIGQCTIYQYDNYNRVIQITLPDGTIVNRKYAPFSTDLITDIRV---NGIS--- 875
             :|| :  |||:|:|||::|::::||||::|||:|||  : |:  :  |   ||:|
Db    145 TEQRDALQQSTLFSYDNWSRMVSSTLADGSVINRSYAPQSSSELATMLEVVHQNGTRTV 204

Qy    876 LGQQTFDGLSRLTQSODGGRVWAYTYSAGNDQCPSTVITPDGQFIHYQYQPELDDAVLQV 935
           |:|:||||.||:|:: ||||:|||:|:|||:| |:||:||| || |||| |  :|| ||
Db    205 AGTQKFDGLERVTQTKTGDRVEQFNYDAGEMQ-PRSRTTAGLDNINFTYTRALTDQIFSS 263

Qy    936 ASNEITQQFSYNPVTGALLKAVAEGQSLTPIYYPSGRLKMENINDMKKMSYLW------T 989
           :  | ::||.||.::|:  ::| : ::||:|||||||::||||||||||:|:|      |
Db    264 TAPDETAKFDYDKTSARLIEATNPQGTRTRYRYDVHNQLTGETWDNL--LGQAWETRHQSS 321

Qy    990 LRGLENGYTDL-------TGTIQKISRDTHGRVTQIKDSSIKTTLNYDDLNRHIGSQVTDLA 1044
           | | ||||||       ||:|:|||||||||:::::::|||||||||||||||||||||:|
Db    322 LLGRPIKRTDLKKGEAAGAETRYRYDYDTLGRIRRFINQSNLRTTTIDYDVLGQLCKVATEDLQ 381

Qy   1045 TGHMLTTTVEFDGLNREIGRKLCDSSGHTLDIQQSWLKTQQLANRIVKLNGVLQRTEQYS 1104
           :|:::   ||||||||:|||||||||||||||||||||||||||:||||||||||||||
Db    382 AGTGVIIDMEYDDQGQEILRTQTASNQAALTLTQTWAVDGLLKTRDLQQAGSPLLHETFS 441

Qy   1105 YDSRNRLNQYKDCGAECPTDKYGHSIVTQNFTYDIYGNITACHTTFADGTEDHATFKFAN 1164
           ||.:|||.:| | |:| |||:  :|:  ::|::||::|||||||||||||:|:|||||.:
Db    442 YDPRGRLTLVNYLGSSLPRDELQREMTRQIFSFDELDNILLCQTRFTDGTSERAAFKYGS 501

Qy   1165 P-----TDPCQLTEVHHTHPDMPDNIRLKYDKAGRVINITDNHGNTENFTYDTLGRL--- 1216
                 |||||| ::  ::  |: :|::|||:.||:|:|::::::| :| |||:::||:
Db    502 PGDDKHKDRCQLLSIAYTPPRKTPDTFSYDANGN--QLKDEHGN--SLHYDSQSRLLQV 557

Qy   1217 -QNGQG--SVVGYDPLNRLVSQKTDLDCEL--YYRETMLVNEVRNGEMIRLLRTGETII 1271
            :||:|  |..|||.:||||: : ::  :|  |:::|:||||::||||||||||||:||
Db    558 AETGGAPISQYRYDGHNQLVATR-DGNESEILRFYEGHQLSSTVQEDRQTQYLHLGBQPL 616
```

```
Qy    474 -----------------------TPYDSEFKDDPEKFIQYRY 511
                                  ||||: |||| |::|||||:
Db    454 DE------HGFVRNLKTRTVQPAATGHGAAATLTQAFQYRALTPLGS------YLKQPW 500

Qy    512 SLIGSQSHVTLKIEERHYSATQLLN--STLFQ-YNTDKSELGRLLKQTECTKG-ENGKTY 567
          :|  | |.|| |:|:|||||||:||  |||:| |||| |::||||||:|:|.| |||:||
Db    501 RLLESET-----LSESGSAPGTFLEKISKLYQEESAERFSYGRVRQQTVSGYPGVDGGSPF 555

Qy    568 SVVHKFTYTKQDD--TLQQSHSITTHDNFTIHRSQVRSRYTGRLFSDTDTKDIVTQMSYD 625
          |:: ::||:|:||  |||||||||||||:|:|||||||||||||||||||||||||:|||
Db    556 DTLTHCGYSLPDDERALQTVEKLVGVDGESKITLRHALDTGEPLLNLDDNGVEIRYYYD 615

Qy    626 KLGRLLTRTLNSGTPYANTLTYDYELNNLQDDNRPPFVITTTDVNGNQLRNEFDGAGRHV 685
          :| |::::|::  :|  ::|||||||||||||||||||||||||||||||||||||||||
Db    616 ALRRVTREVVAPGKPFBAERNVKYFLCAYDNEQAQQWAYDVKQVETHTL--LDGLSRPI 672

Qy    686 SQCLKDSDG---DGKFYTIHTQQYDEQGRHHTSTYSDYLTNGRQQTDPDKVHLSMSKSYD 742
          |:||||||.   |||||||||||||||||||||||||||||||||||||||||||||||
Db    673 FBERADKDSATYAGALRPIYRARYDELDQLVEETEIDWLGDG------LLBLTSHISYD 725

Qy    743 NWGQ---IANTHWSYGVSEKITVDPITLT------ATKQLQSNSNNVQTGKEVTTYTPSQ 793
          :||:   ||| ::|||:::||:| :::|:      | ::: |:|  | .|||:: ::|:
Db    726 DWGQRYAVLNPDGTTQVEE---IDQVASTDGPVHRAWREVE---HSRVSGITETWYNLFE 779

Qy    794 QPIQITLFDEAGHLQ-SCHTLTRDGWDRVRKETDAIG---QCTIYQYDNYNRVIQITLPD 849
          :|::|  |:: |:|: | :|||:||:||||:::::  :   | :::| |:::||:::::|
Db    780 KPVRIERFALDGSTSISLQVNDYDGLGRLSREAQGSGAGQRVVEYRYDAFDRVRRERLAD 839

Qy    850 G-TIVNRKYAPFSDTLITDIRVN-----GISLGQQTFDGLSRLTQSQDGGRVWAYTYSA 903
          | :|:||:||| : :| :|::::     |:||||||||||||||||||||||||||||:
Db    840 RVNTVRTYAGHSRNDLPVSIKVGNTESAAVLLGQVFDGLERRTVAIITGGREQTFEYDP 899

Qy    904 GNDQCPSTVITPDGQFIHYQYQPEL-DDAVLQVASNEITQQFSYNPVTGALLKAVAEGQS 962
          |:::|| || || :::|:|||:|:| ||.|:||:|:  :::|.||||||||||||||||:
Db    900 GERQ-PHWVKAPDGTRIEYQYRPALGEEPVLRVLSGK-EAKYEYDLKNARLTHCEEPGDD 957

Qy    963 LTPIY------YPSGRLKMEN---INDMKKMSYLWTLRGLENGYTDLTGTIQKISRDTH 1012
          ::|:|      |||||||:||   |||||||||||||||||||||||||||||||||:|
Db    958 ENSGYTLDRSHFLSNGEVKRESRTVDGEAFSMTYDYSFRSRLRAVDVDLVGQTQLYDFDDV 1017

Qy   1013 GRVTQIKDSS-----------IKTTLNYDLNRHIGSQVTDLATGHMLTTT 1052
          ||::::|:||           |||||||||||||||||||||||:||:||
Db   1018 GRLEKTLHAPEPKRTRYRLQRAPARQLLLLESTPGYDRQGRMASITTTDASTGHALATL 1077

Qy   1053 VEFDGLNREIGRKLCDSSGHTLDIQQSWLKTQQLANRIVK-------LNGVLQRTEQYS 1104
          :|:|:||||:| |  |:||  |:::||::|::|: |::::|       | |||:|:||:
Db   1078 LEYDEFDREILRTF-DFGDTVQTLAQDYDEFDCLKSRIIKERPKGSDESQATLLRHETYQ 1136

Qy   1105 YDSRNRLNQYKDCGAECPTDKYGHSIVTQNFTYDIYGNITACHTTFADGTEDHATFKFAN 1164
          ||.:|||.:| | |:| |||:  :|:  ::|::||::|||||||||||||:|:|||||.:
Db   1137 YDRRGRLQIYTCDGPEAPVDPSGQTIARQIFGFDGLDNILSVITYRPDGSWQRTLYEFKN 1196

Qy   1165 PTDPCQ------LTEVHHTHPDMPD--------NIRLKYDKAGRVINITDNHGN 1204
              ||      |||| ::  ||.||        |::|||| |||:|:||:::|
Db   1197 -SDPAQMSRIIPPDALLETYDVDHDLEELFKIVKDPQLAIDLHYDGNGNL--ISDEQGR 1253

Qy   1205 TENFTYDTLGRL------QNGQGSVYGYDPLNRL 1232
          :|||:|||:|||      |||||||.|||||||
Db   1254 V--LTYDGLNRLLRVETPDGERCRYNYDPENIL 1284
```

```
Db   185  EEDGSIELVTYREG--GMPFPYYRQ-PPLPCVTLHSLFPGACQSNITDHYFYSGTNYLGF  241
Qy   328  GSNGIWNKLDNLYGLMTEYNYGSTSRRYKDKEGH------DQIVRIERTYNNYHL       378
Db   242  SEPP----DAHQNRLY------YERLELRELVDDEGYYVQILRQNPDHIVSTRHAFNKNHL  292
Qy   379  LTSECKQQNGYIQTTETAYAIIGHNFDSQPSQ--FQLPK-TKTETWRSADNSYRSEIT     434
Db   293  QVRE-----DLQVQFQAEKVISWEFANASPGEKVFGLPTKITTDYTDLSHPNTERTTTV    346
Qy   435  ETTFDESGNPLTKVIKDKKTQKIISPSTHWEYYPPAG------EVD------NC        476
Db   347  QTLAYNNIGQLTKSI------AVDGVVTEWLYYPDTGGQGLDISLIAEKPLSKDLVTLTC   400
Qy   477  PPEPYGFTRFVKKIIQTPYDSEFKDDP------EKFIQYRY------SLIGS          516
Db   401  PKVSEGY-------MPPVKVEYVHDPAYPKSQQITAYAYQARENPVNQRSVLVPSTVVV    452
Qy   517  QSHVTLKIEERHYSATQLLNSTLFQYNTDKSELGRLLKQTECTKG-------ENGKTY--S 568
Db   453  LTGVTL------DTTTMLPSLM------EGRKNALIEQRVISRSIPDVATENTTAWKES    499
Qy   569  VVHKFTYTKQDDTLQQSHSITTHDNFTIHR---------SQVRSRYTGRLFSDTDTK      616
Db   500  VVQNSWLGMQRNTL--TTSMLYDDNPSVGRTVRAEAQGKIISSRIFSRLSGRPLSET-RD   556
Qy   617  DIVTQMSYDKLGRLLTR---TLNSGTPYANTL-TYDYEL--NNLQDDNRPPVITTTDVN    670
Db   557  GLEFRYVHDSLGRLVRQERGTEAGAWKADAVETTDYSITAEGLQ------VTVTEAB      608
Qy   671  GNQLRNEFDGAGRHVSQCLKDSD-GDGKFYTIHTQQYDEQGRHHTSYSDYLTNGRQQTD    729
Db   609  -QQVRTLYDGLQRPVWVAIKRTILPDSAFCVISRIEYDGPDATNQTLY-DYLPGGLRRTK   666
Qy   730  PDKVHLSMSKSYDNWGQIANTHWSYG--VSEKITVDPIPITLTATKQLQSNSNNVQTGKEVT 787
Db   667  DARPEAAVDASKLAW-MADYTREDAGILINEQVIGADSGAQLIRQLSGRNLDKSHTALLE   725
Qy   788  TYTPS--------QQPIQITLPDEAGHLQSCHTLTRDGWDRVRKETAIGQCTIYQYDNYN   840
Db   726  TLRPSTARDASTDRTIERT-FDDERRLIKIRT------SNTSEHC--IEYDELE         770
Qy   841  RVIQITLPDGTIVNRKYAPFSTDTLITDIRVNGISLGQQTFDGLSRLTQSDQGGRVWAYT   900
Db   771  RAVAIIAPDGTRTERKYHQLS--DYITQLNVGSTVLGTQKMTAAARQTTTVGE---LTYE   824
Qy   901  YSAGNDQCPSTVITPDGQFIHYQYQPELDDAVLQVASNEITQQ------FSYNPVT-      950
Db   825  FPGGS---ASTVVRPDKTLLESAVSADGHTATLSINKKVHTQRVISQPNVLTVTVDPVSV   881
Qy   951  -----GALLKAVAEGQSLTPIYY--PSGRLKMENINDMKKMSYLWTGLENGYTDLTGT    1003
Db   882  PSAEAWSSLTSSPQSLGLTSTTQTSPRGSRQAEMTRSLKG------RLLTN--TAVDGR    932
Qy  1004  IQKISRDTHGRVTQIKDSSIKTTLNYDDLNRHIGS-----QVTDLATGHMLTTVEFDGLN  1059
Db   933  QMRVFRDYLDRVVRV----ILGELHYHYLWSAFGEPLQRTVNQASGERLDVRFTWDAFG    988
Qy  1060  REIGRKLCDSSGHTLDIQQSWLKTQQLANRIVKLNGVLQRTEQYSYDSRNRLNQYKC--D  1117
Db   989  QEIAREYTLNNKPLLALTNSYLANGQVSSKTLTREGVLQRTEGFSYDARDRLNSYECTTD  1048
Qy  1118  GAECPTDKYGHSIVTQNFTYDIYGNITACHTTFADGTEDHATFKF---ANPTDPCQL-TE  1173
Db  1049  VADWPQDQAGKSLSKQSYGYDELHNLSECSSTYADGSTCIQTYTYTDTVKNPTRRLSVKTE 1108
Qy  1174  VHH-THPDMPDNIRLKYDKAGRVINITDNHGNTENFTYDTLGRL--QNGQGSV---YGY   1226
Db  1109  LRSGSQTTSTQTATLAYDANGN--QTTDESGRT--LAYTPLGQLASVKDNDGKLLTRYSY  1164
Qy  1227  DPLNRLVSQKTDTL--DCELYYRETMLVNEV---RNGEMRLLRT------GETII       1271
Db  1165  DAFGRLISQYIGATKHTCELLYDGTQLTGEAWFDDANREFKRILFSEDMVQQTCIGETV-  1223
```

```
Qy  1272  AQQ--RASKVLLTGTDSQQSVILTSDKQNLSQEAYSAYGHKSTANDASILGYNGERAD   1328
Db   617  GQQIVDDAEQTLLLLTDANQSVMGEFQQGQLRKAVYSAYGERHSEEALLSTAGFNGEVRE  676
Qy  1329  PVSGVTHLGNGYRSYDPTLMRFHTPDSLSPFGAGGINPYSYCLGDPINRSDPSGH----  1383
Db   677  AANGWYLLGNGYRAYNPLLMRFHSPDFLSPFAEGGVNPYTYCLGNPIALRDPTGHDASGQ  736
Qy  1384  ------LSWQ------AWTGIGMGIA----GLLLTIATGGMAIAAAGGI           1416
Db   737  TGRLRRPDEGALPMQQGGGDIMGWVGVGVVFFTVLGVAATIATLGTATPVTGPVTVLGI   796
Qy  1417  ----AAAIASTSTTALAFG---ALSVTSDITSIVSGALEDASPKASSILGWVS----   1462
Db   797  SMTASAAAAVSTVSTGALIVGTALTAASTTANTVAIVNN----DQTAGEVGGWLGIAAV  851
Qy  1463  ---MGMGAAGLAESAIKGGTKLA-----------THLGAFAEDGENALLKSTSESSR   1505
Db   852  PVGLVGFGAGAVVARAVARAVAAAAKVAAANAGTIGVRSVSVSRIG-LAAAGARRTISSAASSAR 909
```

RESULT 12
Q87VG6

```
ID   Q87VG6      PRELIMINARY;      PRT;   1669 AA.
AC   Q87VG6;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   YD repeat protein.
GN   OrderedLocusNames=PSPTO4970;
OS   Pseudomonas syringae (pv. tomato);
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC   Pseudomonadaceae; Pseudomonas.
OX   NCBI_TaxID=323;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DC3000;
RX   MEDLINE=22834015; PubMed=12928499; DOI=10.1073/pnas.1731982100;
RA   Buell C.R., Joardar V., Lindeberg M., Selengut J., Paulsen I.T.,
RA   Gwinn M.L., Dodson R.J., DeBoy R.T., Durkin A.S., Kolonay J.F.,
RA   Madupu R., Daugherty S.C., Brinkac L.M., Beanan M.J., Haft D.H.,
RA   Nelson W.C., Davidsen T.M., Zafar N., Zhou L., Liu J., Yuan Q.,
RA   Khouri H.M., Fedorova N.B., Tran B., Russell D., Berry K.J.,
RA   Utterback T.R., Van Aken S.E., Feldblyum T.V., D'Ascenzo M.,
RA   Deng W.-L., Ramos A.R., Alfano J.R., Cartinhour S., Chatterjee A.K.,
RA   Delaney T.P., Lazarowitz S.G., Martin G.B., Schneider D.J., Tang X.,
RA   Bender C.L., White O., Fraser C.M., Collmer A.;
RT   "The complete genome sequence of the Arabidopsis and tomato pathogen
RT   Pseudomonas syringae pv. tomato DC3000.";
RL   Proc. Natl. Acad. Sci. U.S.A. 100:10181-10186(2003).
DR   EMBL; AE016874; AAO58398.1; -.
DR   TIGR; PSPTO4970; -.
DR   InterPro; IPR006530; YD.
DR   Pfam; PF05593; RHS_repeat; 5.
DR   TIGRFAMs; TIGR01643; YD_repeat_2x; 2.
KW   Complete proteome.
SQ   SEQUENCE 1669 AA;  184268 MW;  BC668B6C1E3FB7EE CRC64;
```

```
Query Match         10.4%;   Score 920.5;  DB 2;  Length 1669;
Best Local Similarity   25.9%;   Pred. No. 8e-40;
Matches   440;  Conservative 212;  Mismatches 662;  Indels 387;  Gaps 84;
```

```
Qy   152  AFDLKVPKKLLNPAGHAIYIDWNFEATQPRLNRIYDDLDGHDIPLLNLEYQGL---IKTI  208
Db    72  AFGQLLPSEVTSPQGGTLSIEWSRNQGQFLKKI--SSDGNVI--LSSTYSAPSSDTSTV   127
Qy   209  -LTLFPGQKEGYRTELRFLNRQLNSIHNFSLGNENPLTWSFGYTPIGKNGILGQWITSMT  267
Db   128  EISVWPKTDEAYTVRLELSSYALTGLTRIEKNLKVSQKLKYGYC--ADPTLDRVLNRIE   184
Qy   268  APGGLKETVNYSNNNQGHHFPQSANLPLPYVTLMKQVPGAGQPAIQAEYSYTSHNYVGG   327
```

```
Qy   196 LLNLEYQGLIKTILTLFPGQKEGYRTELRPLNRQLNSIHNFSLGNENPLTWSFGYTPIGK 255
             :|| |     |||| | :|::
Db   173 LCKIDYQFSDWARITFWPGKTESYTFQLDFVNEYLYWVTNKSTSRE--LVWSFNYDDVGA 230

Qy   256 NGILGQWITSMTAPGGLKETVNYSNNNQGHHFPQOSANLPVLPYVTLMKQVPGAGQPAIQA 315
             |||  ||:        ::|||||
Db   231 GNFT---LTQVKSPTGLTETVNYQAGVM--RFPDESGKPALPSVNYRQSPGMGQPDIVK   285

Qy   316 EYSYTSHNYVG-GGSNGIWNNKLDNLYG-LMTEYNYGSTESRRYKDKEGHDQIVRIERTY 373
             ||||||||| | :  ||  :|::       |||
Db   286 EYEYTASNYLJGYGASLGKANEDEDNIYNVMDDYYYSSTEKLIVDNRE---LVSISRIY   341

Qy   374 NNYHLLTSECKQONGYIQTTETAYYAIIGHNFDSQPSQFQLPKTKTETWR-SADNSYRSE 432
             :|:||:|     |::||||
Db   342 NSYYLLISETTRQNSCEVIVETDYYAKPGLSFDKQPKQFQLPKEEKKTWRENSKNQCRSE 401

Qy   433 ITETTFDESGNPLTKVIKDKKTQKIISPSTHWEYYPPAGEVDN---CPPEPYGFTRFVK 488
             |::|| :| ::  :|:||
Db   402 ITTTFDPEGNLLTKIEPD--------GTKTEYIIYDSKGETDKGIVLCPPEPNGFVRFVK 454

Qy   489 KIIQTPYDSEFKDDPEKFIQYRYS----LIGSQSHVTLKIEERHYSAT 532
             ||:|       || :|     ||::
Db   455 TQIVTPANSEFY-APVQQTYAYAQYPCIAGSSLSYAVLQNKKHFAVT 501


RESULT 14
Q9X626
ID   Q9X626                  PRELIMINARY;    PRT;    526 AA.
AC   Q9X626;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
OS   Coxiella burnetii.
OG   Plasmid QpDV.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
OC   Coxiellaceae; Coxiella.
OX   NCBI_TaxID=777;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Radomski K.U., Willems H., Lautenschlaeger S., Jaeger C., Baljer G.;
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF131076; AAD33498.1; -.
KW   Hypothetical protein; Plasmid.
SQ   SEQUENCE 526 AA;  60111 MW;  AF26E1D9E1B56834 CRC64;

Query Match             10.4%;  Score 919.5;  DB 2;  Length 526;
Best Local Similarity   40.8%;  Pred. No. 1.5e-40;
Matches  221;  Conservative  73;  Mismatches 203;  Indels  45;  Gaps  16;

Qy    18 NEF-FTQANNFTSAVSGGVDPRTGLYNIQITLGHIVGNGNLGPTLPLTLSYSPLNKTDI- 75
             :| :|| || ||::|:|||||||||
Db     2 NELPYTQATNFISAVQGGVDPRTGLFTVNMVLAELTGNDNLGPDLFTLNYSPLSTSNIC 61

Qy    76 GFGIGENFGLSVYDRKNSLLSLSTGENYKVIETDKTVKLQQKKLDNLRFEKDLKENCYRI 135
             ||||| ||| || :||||| ||:|| ||:  ::|||
Db    62 GFGIGCSVGISIYDKNNKLLILSSSERYKTEDWNDSVYVRQKKINNFKFEK--IKNGYII 119

Qy   136 IHKSGDIEVLTGFNNNAFDLKVPKKLLNPAGHAIYIDWNFEATQPRLNRIYDDLDGHDIP 195
             :|:      |||   |   ::|||
Db   120 KYKNGKTEYLYKYGDNLF--LPQKIFSTLGWPLKLTWENRGQYVNLTKIEDAKD----V 172

Qy   196 LLNLEYQGLIKTILTLFPGQKEGYRTELRPLNRQLNSIHNFSLGNENPLTWSFGYTPIGK 255
             :|| |     |||| | :|::
Db   173 LCKIDYQFSDWARITFWPGKTESYTFQLDFVNEYLYWVTNKSTSRE--LVWSFNYDDVGA 230

Qy   256 NGILGQWITSMTAPGGLKETVNYSNNNQGHHFPQOSANLPVLPYVTLMKQVPGAGQPAIQA 315
             |||  ||:        ::|||||
Db   231 GNFT---LTQVKSPTGLTETVNYQAGVM--RFPDESGKPALPSVNYRQSPGMGQPDIVK   285

Qy   316 EYSYTSHNYVG-GGSNGIWNNKLDNLYG-LMTEYNYGSTESRRYKDKEGHDQIVRIERT 372
             ||||||||| |  ||  :|::      |||
Db
```

```
Qy   1272 AQQRASKVLLTGTDSQOSVILTSDKQ-----NLSQRAYSAYGKHKSTANDASI-IGYNGE 1325
               :::|         :::   : :|:  ||: ||:
Db   1224 ---RSDFVL---TDPGSGGVGFSAHDGTAGVKHPLGYTPYG--ESTNLDSGCRLGFNSE 1275

Qy   1326 RADPVSGVTHLGNGYRSYRYDPTLMRFHTPDSLSPFGAGGINPYSYCLGDPINRSDPSGHL- 1384
               ||||| |||| |||||| ::| |
Db   1276 RIDPVLGWYHLGNGYRYSPAQRHWLQPDSWSPFGAGGINNTAYCAGDPVNLFDPSGHVM 1335

Qy   1385 --SWQA-------------WTGI----GMGIAGLLLTIATGGMAIAAAGGI 1416
               :                    |:     ::|:|||
Db   1336 ISRWGASMISDLTKALQETSPQPLGHFWRGLAVNASVAVAGVLMVPLTCGSSLGFAAGV 1395

Qy   1417 AAAIASTSTTALAFGALSVTSDITSIVSGALEDASPK-------ASSILGWVSMGMGAAG 1469
               :::|     |  :::::           :|
Db   1396 LA------TTLAVASAGL---DIASYV---LEDVNPBLARKIGTAATALGFISNAPFRAG 1443

Qy   1470 L------------AESAIKGGTKLATHLGAF--AEDGENALLKSTSESSRIK 1507
               |            :|  :|:|: :|      :  ||:
Db   1444 LRLGGRLLLRWTSSRTGRLISRVESIFTKGPKLKLLTGTIKVLKPG----MSSYGASNVTR 1499

Qy   1508 WGVTRSLDREIVRNEEGQVIKDHSRGYTDNFMGKGEQAILVHGDKDG-----FLYHTEGNK 1563
               :   :     :::| :|:||
Db   1500 W----------APGKLDKF-EKLENALFMTVDKQGKRVTFMAH--GVK 1534

Qy   1564 ----------HNGKGPYT--RHTPEQLVDYLKDNNIVDLTCGGDKPVHLLSCYGKS 1607
                       ::||::                  |:  |:   :|:| :::
Db   1535 PDDVGDAAMMAIEEYSGSGGPYSGKSPSAYSFFDFLKSKK--VDLNK--YEKVRLIMCHSAD 1591

Qy   1608 SGA---ADKMAKYINRPVIAY 1625
               ::|   ||: |:| |:||
Db   1592 GGEKSFAATFSKLTNKPVKGY 1612


RESULT 13
Q45905
ID   Q45905                  PRELIMINARY;    PRT;    505 AA.
AC   Q45905;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Orf 505 protein.
GN   Name=orf 505;
OS   Coxiella burnetii.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
OC   Coxiellaceae; Coxiella.
OX   NCBI_TaxID=777;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97294470; PubMed=9150226;
RA   Willems H., Ritter M., Jager C., Thiele D.;
RT   "Plasmid-homologous sequences in the chromosome of plasmidless
RT   Coxiella burnetii Scurry Q217.";
RL   J. Bacteriol. 179:3293-3297(1997).
DR   EMBL; X93204; CAA63682.1; -.
SQ   SEQUENCE 505 AA;  57841 MW;  8B60B4C307CA8D0B CRC64;

Query Match             10.4%;  Score 919.5;  DB 2;  Length 505;
Best Local Similarity   41.1%;  Pred. No. 1.4e-40;
Matches  217;  Conservative  75;  Mismatches 195;  Indels  41;  Gaps  16;

Qy    18 NEF-FTQANNFTSAVSGGVDPRTGLYNIQITLGHIVGNGNLGPTLPLTLSYSPLNKTDI- 75
             :| :|| || ||::|:|||||||||
Db     2 NELPYTQATNFISAVQGGVDPRTGLFTVNMVLAELTGNDNLGPDLFTLNYSPLSTSNIC 61

Qy    76 GFGIGENFGLSVYDRKNSLLSLSTGENYKVIETDKTVKLQQKKLDNLRFEKDLKENCYRI 135
             ||||| ||| || :||||| ||:|| ||:  ::|||
Db    62 GFGIGCSVGISIYDKNNKLLILSSSERYKTEDWNDSVYVRQKKINNFKFEK--IKNGYII 119

Qy   136 IHKSGDIEVLTGFNNNAFDLKVPKKLLNPAGHAIYIDWNFEATQPRLNRIYDDLDGHDIP 195
             :|:      |||   |   ::|||
Db   120 KYKNGKTEYLYKYGDNLF--LPQKIFSPLGWPLKLSWENRGQYVNLTKIEDAKD----V 172
```

```
Db   286  EYEYTASNYLGYGASLGKAWNEBEDNIYNVVMDDYTYSSTEKLIVDNRE----LVSISRI  341    Qy   432  EITETTFDESGNPLTKVIKDKKTQKIISPSTHWEYYPPAGEVDN----CPPEPYGFTRFV  487
                                                                                    Db   402  EITTTTFDPEGNLLTKIEPD------GTKTEYIYDSKGETDKGIVLCPPEPNGFVRFV  454
Qy   373  YNNYHLLTSECKQQNGYIQTETAYYAIIGHNFDSQPSQFQLPKTKTETWR-SADNSYRS  431
                                                                                    Qy   488  KKIIQTPYDSEF----KDDPEKFIQYRYSLIGSQSHVTLKIEBRHYSATQLLNSTLFQYN  543
Qy   342  YNSYYLLISETTRQNSCEVIVETDYYAKPGLSFDKQPKQFQLPKEEKKTWRENSKNQCRS  401    Db   455  KTQIVTPADSEFYAPVQQTTYAYAQYPCIAGSSLSYAVLQTQETLCSDDVL----LLTIN  510

Qy   432  EITETTFDESGNPLTKVIKDKKTQKIISPSTHWEYYPPAGEVDN----CPPEPYGFTRFV  487    Qy   544  TD  545
Db   402  EITTTTFDPEGNLLTKIEPD------GTKTEYIYDSKGETDKGIVLCPPEPNGFVRFV  454    Db   511  TD  512

Qy   488  KKIIQTPYDSEF----KDDPEKFIQYRYSLIGSQSHVTLKIEBRHYSATQLLNSTLFQYN  543
Db   455  KTQIVTPADSEFYAPVQQTTYAYAQYPCIAGSSLSYAVLQTQETLCSDDVL----LLTIN  510    Search completed: February  9, 2005, 18:13:59
                                                                                    Job time : 240 secs
Qy   544  TD  545
Db   511  TD  512


RESULT 15
O52883
ID   O52883        PRELIMINARY;    PRT;    526 AA.
AC   O52883;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein orf 526.
GN   Name=orf 526;
OS   Coxiella burnetii.
OG   Plasmid QpRS.
OC   Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
OC   Coxiellaceae; Coxiella.
OX   NCBI_TaxID=777;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Lautenschlaeger S., Jaeger C., Willems H., Baljer G.;
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Y15898; CAA75844.1; -.
KW   Hypothetical protein; Plasmid.
SQ   SEQUENCE   526 AA;  60104 MW;  42D04FF8CCC361CC CRC64;

Query Match           10.3%;   Score 916.5;  DB 2;  Length 526;
Best Local Similarity 40.8%;   Pred. No. 2.1e-40;
Matches  221;  Conservative  73;  Mismatches 203;  Indels  45;  Gaps  16;

Qy   18   NEF-FTQANNFTSAVSGGVDPRTGLYNIQITLGHIVGNGLGPTLPLTLSYSPLNKTDI-  75
               ::|||:|||:||:||||:||||||| ||:||||||||:|||||||||||:||:|:
Db   2    NELPYTQATNFISAVQGVVDPRTGLFTVNMVLAELTGNDNLGPDFLFTLNYSPLSTSNIC  61

Qy   76   GFGIGFNFGLSVYDRKNSLLSLSTGENYKVIETDKTVKLQQKLDNLRFEKDLKENCYRI  135
          ||||||||||||| |||| ||||||||||:||||||:|:||||||||:|::|| |||||:
Db   62   GFGIGCSVGISIYDKNKLLLLSSGERYKIEDWNDGVYVRQKKINNFKFEK--IKNGYII  119

Qy   136  IHKSGDIEVLTGFNNNAFDLKVPKKLLNPAGHAIYIDWNFEATQPRLNRIYDDLGDGHDIP  195
          :|::|| | |:| |  || |:|| :||||:|||  |:||  |  |
Db   120  KYKNGKTEYLYKYGDNLF---LPQKIFSTLGWPLKLTWENRGQYVNLKKIEDAKD----V  172

Qy   196  LLNLEYQGLIKTILTLFPGQKEGYRTELRFLNRQLNSIHNFSLGNENPLTWSFGYTPIGK  255
             : |:||| | |||:|:| |||:|  ||:| |::|||||:  |||||||:
Db   173  LCKIDYQFSDWARITFWPGKTESYTFQLDFVNEYLYWVTNKSTSRE--LAWSFNYDDVGA  230

Qy   256  NGILGQWITSMTAPGGLKETVNYSNNNQGHHPPQSANLPVLPYVTLMKQVPGAGQPAIQA  315
          ||  |  :|||:  |||||||||| ||   | :||||||||:||||||:|::|
Db   231  GNFT---LTQVKSPTGLTETVNYQAGVM--RFPDESGKPALPSVYNYRQSPGMGQPDIVK  285

Qy   316  EYSYTSHNYVGGGSN--GIWNNKLDNLYG-LMTEYNYGSTESRRYKDKEGHDQIVRIERT  372
          |||||:|||| ||||  |||||:||||:|  |:|||||||:::|| |:|:||||::
Db   286  EYEYTASNYLGYGASLGKAWNEBEDEDNIYNVVMDDYTYSSTEKLIVDNRE----LVSISRI  341

Qy   373  YNNYHLLTSECKQQNGYIQTETAYYAIIGHNFDSQPSQFQLPKTKTETWR-SADNSYRS  431
          |||:||||||| | ||||||:|||:|||:|||||||||:||| :|:||| |::|| ||||
Db   342  YNSYYLLISETTRQNSCEVIVETDYYAKPGLSFDKQPKQFQLPKEEKKTWRENSKNQCRS  401
```

```
              GenCore version 5.1.6
          Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on:     February  9, 2005, 17:59:43 ; Search time 58 Seconds
                                (without alignments)
                              2775.356 Million cell updates/sec

Title:      US-09-889-874A-23
Perfect score: 8879
Sequence:   1 VYIKFLKLFRRITMSDNNEF.........PRKIILGRTEKTVKPKTFRP 1673

Scoring table: BLOSUM62
            Gapop 10.0 , Gapext 0.5

Database :  PIR_79:*
            1: pir1:*
            2: pir2:*
            3: pir3:*
            4: pir4:*

Searched:   283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters:    283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                         SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	Description				
1	1214	13.7	709	2	S38241	hypothetical prote	30	245	2.8	2825 2 T14271	Doc4 protein, stre
2	912.5	10.3	528	2	S38242	hypothetical prote	31	230	2.6	794 2 T36972	probable membrane
3	432	4.9	2334	2	S32920	cell wall-associat	32	230	2.6	849 2 T46253	hypothetical prote
4	429.5	4.8	336	2	S38239	hypothetical prote	33	220.5	2.5	1185 2 A42404	collagen adhesin -
5	386	4.3	2167	2	AF1489	cell wall-associat	34	219.5	2.5	1962 2 A32634	lactocepin (EC 3.4
6	381	4.3	2183	2	T37218	hypothetical prote	35	218.5	2.5	2406 2 A54148	odz protein - frui
7	365	4.1	1400	2	E90886	RhsE core protein	36	217	2.4	356 2 T37136	hypothetical prote
8	353.5	4.0	843	2	AB0539	Rhs-family protein	37	216.5	2.4	2894 2 C64474	hypothetical prote
9	342	3.9	1397	2	C64805	rhsC protein precu	38	215.5	2.4	2044 2 AB1180	probable peptidogl
10	341.5	3.8	1404	2	B90658	RhsG core protein	39	213	2.4	1487 2 AG2560	hypothetical prote
11	341	3.8	1394	2	H91236	RhsH core protein	40	211	2.4	3283 2 AC1018	large repetitive p
12	340.5	3.8	1411	2	B65145	rhsB protein precu	41	208	2.3	2703 2 H81193	hemagglutinin/hemo
13	339	3.8	1426	2	H64780	rhsD protein precu	42	207	2.3	2893 2 A64556	toxin-like outer m
14	336.5	3.8	1404	2	B85509	hypothetical prote	43	206	2.3	1902 2 S06997	lactocepin (EC 3.4
15	331.5	3.7	1377	2	C65159	rhsA protein precu	44	203.5	2.3	1366 2 S57664	IgA-specific metal
16	331	3.7	1409	2	F91187	rhsA core protein	45	202.5	2.3	1349 2 AI1476	cell surface prote
17	326	3.7	1399	2	A99720	RhsC core protein					
18	322.5	3.6	1377	2	E86034	rhsA protein in rh					
19	319	3.6	1397	2	A85570	rhsC protein in rh					
20	316	3.6	1398	2	B85549	hypothetical prote					
21	313	3.5	1398	2	H90698	RhsD core protein					
22	297	3.3	1512	2	AH0439	probable membrane					
23	280	3.2	1317	2	F83310	conserved hypothet					
24	278	3.1	985	2	B86084	hypothetical prote					
25	276.5	3.1	1438	2	AI0093	conserved hypothet					
26	271.5	3.1	1354	2	AG0538	Rhs-family protein					
27	266.5	3.0	2515	2	S47008	tenascin-like prot					
28	248.5	2.8	1616	2	E90704	Rhs core protein w					
29	245.5	2.8	1645	2	H85554	hypothetical prote					

```
                         ALIGNMENTS

RESULT 1
S38241
hypothetical protein - Coxiella burnetii
C;Species: Coxiella burnetii
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Jul-2004
C;Accession: S38241
R;Thiele, D.; Willems, H.; Haas, M.; Krauss, H.
submitted to the EMBL Data Library, October 1993
A;Reference number: S38215
A;Accession: S38241
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-709 <THI>
A;Cross-references: UNIPROT:O45948; EMBL:X75356; NID:g407370; PIDN:CAA53129.1; PID:g4073

Query Match        13.7%;  Score 1214;  DB 2;  Length 709;
Best Local Similarity   40.8%;  Pred. No. 5.6e-59;
Matches 287;  Conservative 101;  Mismatches 280;  Indels 36;  Gaps  16;

Qy   554 KQTECTKGENGKTYSVVHKFTYTKQDDTLQQSHSITTHDNFTIHRSQVRSRYTGRLFSDT 613
          ::   :  : |:  :: ||    ::.:| .:|.|.::  ::||..||.|:|:|:||.|.
Db     3 RRAEVLTSKEGKKYQNQTTFAISSQAEHLLQKIDFTGGDGIKTSISREQSRYSGRLLSST 62

Qy   614 DTKDIVTQMSYDKLGRLLTRTLN-SGTPYANTLTYDYELNNLQDDNRPPFVITTTDVNGN 672
          |||  ||: ||.|.||::|.: |  |: ||| :.|.  :||:|::|.||: |::||||
Db    63 DELGNVTQYEYDELGRLLLTQTVNASSTTYASTTRYSYSLETDARGKVTAIKTTVTDPKGN 122

Qy   673 QLRNEFDGAGRHVSQCLKDSDG------DGKFYTIHTQQYDEQGRHHTSTYSDYLTNGRQQ 727
          |||:  ||.|| || ::|:|    ||.|.|.||.| ||.|. : |   :|:  :|:|::|
Db   123 QLRTYYDGLGRNLKQERLDKDAAVSQKTGTWYITHQQQYDALGRESKTITQDVL----R 177

Qy   728 TDPDKVH-----LSMSK--SYDNWGQIANTHWSYGVSEKITVDPITLTATKQLQSNSNNVQ 781
          ||| :|     ||:||  |.||.|| |.|||||.|.:. |:||:|||:|.||:|.|.:|:|
Db   178 LDSEVGHAGSVLSTSKVMHYDSWGQNHLTVFSDGYQERSVYDPITRRAT--LQPESGSQK 235

Qy   782 TGKEVTYTPSQQPIQITLFPEAGHLQSCHTLTRDGWDRVRKETDAIGQCTIYQYDNYNR 841
          ||.|.|| |.|| :|:.|:|.   ::|.|.||:||.|  :| ::|| ::||.|||.||:|
Db   236 LGQQLTEYNLAGLPIKVTQYDSQGTEQGSAHYEYDGLGQLRKETDELGQITLYEYDHFGR 295

Qy   842 VIQITLPDGTIVNRKYAPFSTDTLITDIRVNGISLGQQTFDGLSRLTQSQDGGRVWAYTV 901
          |:| |:|||.:||.| || ||| |.|:. ::|:|.||.||:||.||||| ||.|.|.:
Db   296 VTQTTLPENTIIQKSYAPHSTASLITGISVNNFSMGNQTFDSLERLTETTSGGRTSAFSY 355

Qy   902 SAGNDQCPSTVITPDGQFIHYQYQPELDDAVLQVASNEITQQFSYNPVTGALLKAV-AEG 960
          |:|:|.| |:|:||||.|:||.|:|.|:|:|::| :||:||:|||:|.|| |:||. |.|
Db   356 ENAS-SVPAAVTAPTGETVSYEYLKELGNAVKKISAPEILQTWDYDALTGAMTSATQAAG 414

Qy   961 QSLTPIYPSGRLKME-NIND---MKKMSYLWTLRGLENGYTDLTGTIQKISRDTHGRVT 1016
          ||:||||:|.|.|:  ||:    :  :| :|  ||||::|:|||.| :|:|.|||.|.|
Db   415 MIRQMTYYPSGLLKNETSMPDGAAQKSTAYTYSLAGAPQSYTDVFGVTQRQYDEHGRRI 474

Qy  1017 QIKDSSIKTLNYDDLNRHIGSQVTDLATGHMLTTVEFDGLNREIGRKLCDSSGHTLDI 1076
```

```
Db      475  GIEENDIKVSLDYDAFGRFTKQQATDKKTGAVLSTTLTYDDLNREIKREISASGQSVLVI  534
             |::|:|| :|:||   |  |   || |  ||||  |:::  ||||| |||||||
Qy      1077 QQSWLKTQQLANRIVKLNGVLQRTEQYSYDSRNRLNQYKCDGAECPTDKYGHSIVTQNFT  1136
             :|:::  |  || || :| |  ||||| ::  | |  ||| :|: ::|
Db      535  EQTYQRNHLLKERITQRGRTTLRKEMFAYDSRNRLIEYTCNGEARPQDPYGKAIHRQTFS  594
             |  |||:  | |||||  | | |||  ||| | ||:|  :|||   ||| |:  |
Qy      1137 YDIYGNITACHTTFADGTEDHATFKPANPTDPCQLTEVHHTHPDMPDNIRLKYDKAGRVI  1196
             :| ||:  |   :: |  ||  ::  |||  :|||| ||:||||: :| | :| |||  :
Db      595  YDALGNMTKTQTDFSGG-RNTATYIY-SAIDPTQLLKVNNDHSDYPKEITLEYDKAGRMI  652
             |  ||| :  |: |   |  |  :|: |:|  |  :::|  :|| |   |  ||||| :
Qy      1197 NITDNHGNTENFTYDTLGRLQ--NG--QGSVVGYDPLNRLVSQ  1235
             |  |:|||:| | |||  ||:   ||  ||:| | |||  |||
Db      653  R--DEAGRT--LRYDALGRLQQVNGAGAKGGQYAYDVLNTLVSQ  692
```

RESULT 2
S38242
hypothetical protein - Coxiella burnetii
C;Species: Coxiella burnetii
C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Jul-2004
C;Accession: S38242
R;Thiele, D.; Willems, H.; Haas, M.; Krauss, H.
submitted to the EMBL Data Library, October 1993
A;Reference number: S38215
A;Accession: S38242
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-528 <THI>
A;Cross-references: UNIPROT:Q45949; EMBL:X75356; NID:g407370; PIDN:CAA53130.1; PID:g4073
C;Superfamily: Coxiella burnetii hypothetical protein

Query Match   10.3%; Score 912.5; DB 2; Length 528;
Best Local Similarity 40.8%; Pred. No. 1.4e-42;
Matches 222; Conservative 72; Mismatches 203; Indels 47; Gaps 17;

```
Qy      18   NEF-FTQANNFTSAVSGGVDPRTGLYNIQITLGHIVGNGNLGPTLPLTLSYSPLNKTDI-  75
             ||: || | | ||  | ||||||||| ::: : : :  |  |||:|| |:||||||:|
Db      2    NELPYTQATNFISAVQGGVDPRTGLFTVNMVLAELTGNDNLGPDFLFTLNYSHLSTSNIC  61

Qy      76   GFGIGFNFGLSVVDRKNSLLSLSTGENYKVIETDKTVKLQQKKLDNLRFEKDLKENCYRI  135
             ||||| :|| ||  ||:  ::  ||||:|||||||||||| ||||:|||:: | |:|||
Db      62   GFGIGCSVGISIYDKNNKLLILSSGERYKTEDWNDSVVRQKKINNFKPEK--IKNGYII  119

Qy      136  IHKSGDIEVLTG--FNNNAFDLKVPKKLLNPAGHAIYIDWNFEATQPRLNRIYDDLLDGHD  193
             | :|||:||| |  |||  ||| ||  |||||||||||||| |||||||||||||||| |
Db      120  KYKNGKTEYLYNYKYGDNLF---LPQKIFSPLGWPLKLTWENRGQYVNLTKIEDAKD---  173

Qy      194  IPLLNLEYQGLIKTILTLFPGQKEGYRTELRFLNRQLNSIHNFSLGNENPLTWSFGYTPI  253
             ||||||:||| ||| |   |   : | :  |:|| ||  |:||  |||||| |||| |:
Db      174  -VLCKIDYQFSDWARITFWPGKTESYIFQLDFVNEYLYWVTNKSTSRE--LVWSFNYDDV  230

Qy      254  GKNGILGQWITSMTAPGGLKETVNYSNNNQGHHPPQSANLPVLPYVTLMKQVPGAGQPAI  313
             |||: |||| |::||  || |:||| ::|: | | : ||||  || :| ||| ||| |
Db      231  GAGNFT---LIQVKSPTGLTETVNYQAGVM--RFPDESGKPALPSVNYRQSPGMGQPDI  285

Qy      314  QAEYSYTSHNYVGGGSN--GIWNNKLDNLYG-LMTEYNYGSTESRRYKDKEGHDQIVRIE  370
             |||||:|| | |  |:|  ||||| || |  |:|||||||||||| |  :| | |:|
Db      286  VKEYEYTVSNYLGYGASLGKAWNEDEDNIYNVMDDYTYSSTEKLIVDNRE----LVSIS  341

Qy      371  RTYNNYHLLTSECKQQNGVIQTTETAYYAIAIGHNFDSQPSQFQLPKTKTETWR-SADNSY  429
             |  |:| |  |: :|:||:  ||| ||  ||::  |||  | || ||   ||  :| ||:
Db      342  RIYNSYYLLISETTRQNSCEVIVETDYYAKPGLSFDKQPKQFQLPKEEKKTWRENSKNQC  401

Qy      430  RSEITETFDESGNPLTKVIKDKKTQKIISPSTHWEYYPPAGEVDN---CPPEPYGFTR  485
             |||||:|||| |||| :|:|||||||::|| |:||| ||||  |   ||| ||| ||
Db      402  RSEITTTTFDPEGNLLITKIEPD------GTKTEYIIYDSKGETDKGIVLCPPEPNGFVR  454

Qy      486  FVKKIIQTPYDSEF---KDDPEKFIQYRYSLIGSQSHVTLKIEERHYSATQLLNSTLFQ  541
             || :||:|||   |   | |||:| ||||| |:|| |||||  ||| ||| |||  |
Db      455  FVKTQIVTPADSEFYAPVQQTTVAYAQYPCIAGSSLSVAVLQTETLCSDDVL----LLT  510

Qy      542  YNTD  545
             :|:|
Db      511  INTD  514
```

RESULT 3
S32920
cell wall-associated protein precursor wapA [similarity] - Bacillus subtilis
C;Species: Bacillus subtilis
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 09-Jul-2004
C;Accession: S32920; E69730; T47101
R;Foster, S.J.
Mol. Microbiol. 8, 299-310, 1993
A;Title: Molecular analysis of three major wall-associated proteins of Bacillus subtilis
rotein.
A;Reference number: S32919; MUID:93302506; PMID:8316082
A;Accession: S32920
A;Molecule type: DNA
A;Residues: 1-2334 <FOS>
A;Cross-references: UNIPROT:Q07833; GB:L05634; NID:g304177; PIDN:AAA22883; PID:g304179
R;Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertram,
A.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galler
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
Y.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A;Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A;Reference number: A69580; MUID:98044033; PMID:9384377
A;Accession: E69730
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-2334 <KUN>
A;Cross-references: GB:Z99124; GB:AL009126; NID:g2636442; PIDN:CAB15959.1; PID:g2636469
A;Experimental source: strain 168
R;Yoshida, K.; Sano, H.; Seki, S.; Oda, M.; Fujimura, M.; Fujita, Y.
Microbiology 141, 337-343, 1995
A;Title: Cloning and sequencing of a 29 kb region of the Bacillus subtilis genome contai
A;Reference number: Z24350; MUID:95219088; PMID:7704263
A;Accession: T47101
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2334 <YOS>
A;Cross-references: EMBL:D31856; NID:g603765; PIDN:BAA06656.1; PID:g603782
A;Experimental source: strain BGSC1A1
C;Genetics:
A;Gene: wapA; N17G
C;Superfamily: cell wall-associated protein wapA

Query Match   4.9%; Score 432; DB 2; Length 2334;
Best Local Similarity 20.4%; Pred. No. 5.6e-15;
Matches 358; Conservative 226; Mismatches 613; Indels 560; Gaps 80;

```
Qy      44   IQTLGHIVG-NGNL--------------GPTLPLTLSYSPLNKTDIGFGIGFNFGLSVVDR  90
             | || |||| | ||              || ||||:|||| ||||||| ||||||:::
Db      896  IDIPSGQLNGATGNVIVNEEDLSIDGRGPGLGLSRTYNSLDSSDHLFGQGW-----YADA  950

Qy      91   KNSLLSLSTGENYKVIETDKTVKLQQKKLDN-----LRFEKDLKENCYRIIHKSGDIEVL  145
             | | :  ||| |:    ||   | :::  ||     |     |:    |: :: | |  |
Db      951  ETSVISTDGGAMY--IDEDATTHRFTKKADGTYQPPTGVYLEJTETADQFILKTKD----  1004

Qy      146  TGFNNNAFDLKVPKKLL------NPAGHAIYIDWN-----FRATQPRLNRIYDDLLDGH-  192
             ||       | : |:         ||  |  | ||     | |  |: |||||  |: |
Db      1005 --QTNAYFNKKGGKLQKVDGHNNATVTVTYNDKNQLTAITDASGRKLTFTYDE-NGHVT  1060

Qy      193  DIPLLNLEY--QGLI-----------------------KTILTLFPG  214
             || || ::   ||||                       |||||  ||
Db                                       :::   :|:
```

```
Qy  1061 SITGPKNKVTYSYENDLLKKVTDTDGTVTSYDYDSEGRLVKQYSANSTEAKPVFTEY--  1118
Db  1175 H------------------------------------------HTHP-DMPDNI---------RLKYDKAGRVINITDNHG  1203

Qy   215 QKEGVRTELRPLNRQLNSIHNFS------LGNENPLTWSFGYTPIGKNGILGQWITSMT   267
Db  1974 NDESISYDKNGNRTSDGKFYTWDAEDNLTAVTKKGEDKPFATYKYDEKGNRIQKTVN-G  2032

Qy  1119 QYSGHRLEKAINAKKETYYSYDADKKTLLMTQPNGRKVQYGYNEAGNP------IQVID  1172
Db  1204 NTENFTYDTLGRLQNGQGSVVGYGYDPLNRLVSQKTDTLDCELY-YRETMLVNEVRNGEMIR  1262

Qy   268 APGGLKETVN--YSNNN--------QGHHFPQSANLPVLPYPVTLMKQVPGAG        309
Db  2033 KVTNYFYDG---------------------DSLNVLYETDADNNVTKSYTYGD----------SGQLLS  2070

Qy  1173 DAEGLKITTNTKYEGNNVVEDVDPNDVGTGKATESYQYDKDGN-----VTSVKDAYGT-  1225
Db  1263 LLRTGETIIAQQRASKVLLTGTDSQQSVILTSDKQNLSQEAYSAYGK-HKSTANDA---S  1318

Qy   310 QPAIQAEYSYTSHNYVGGGSNGIWNNKLDNLYGLMTEYNYGSTESRRYKDKEGHDQIVRI   369
Db  2071 YTENGKKYFYHYNAHGDIIAISDSTGKTV---------AKYQYDAWGNPTKTEASDEVKDN  2122

Qy  1226 ------ETYEYNKNNDV--------TKMKDTEGNVTDIAYDGLDAVSETDQSGKSSSAAV  1271
Db  1319 ILGYNGERADPVSGVTHLGNGYRSYDPTLMRFHT--PDSLSPFGAGGINPYSYCLGDPIN  1376

Qy   370 ERTYNNYHLLTSECKQQNGYIQTETAYYAIIGHNFDSQPSQFQLPKTKTETWRSADNSY   429
Db  2123 RYRYAGYQYDEETGLYYLMARY--YEPRNGVFLSLDPDPGSDGDSLDQNGYAYGNNNPVM  2180

Qy  1272 YDKYGNQIQSSKDLSASTNILK----------DGSFEAQKSGWNLTASKD---------  1311
Db  1377 RSDPSGHLSWQAWTGIGMGIAGLLLTIATGMAIAAAG---GIAAAIASTSTTALAFGA  1432

Qy   430 RSEITETTFDESGNPLTKVILKDKKTQKIISPS----THWEYYPPAGEVDNCPPEPYGFTR   485
Db  2181 NVDFDGH--W-VWLVVNAGPA---------AYDGYKAYKSGKGWKGAAWAAASNFGPGKIFKG  2231

Qy  1312 RRKIS-VIADKSG------VLSGSKALEVLSQSTSAGTDHGYSSATQTVBLEPNTTYTLSG  1365
Db  1433 LSVTSDITSIVSGALEDASPKASSILGWVSMGM----GAAGLAESAIKGGTKLATHLGAF  1488

Qy   486 FVKKII---QTPYDSEFKDDPEKFIQY--RYSLIGSQSHVTLKIEERHYSATQLLNS-   537
Db  2232 ASRAYKFTK--------------------KAVKLTGHTRHGLNQSIGRNG---------GRGVNLRAKLNA-  2273

Qy  1366 KIKTDLAKSRAYFNIDLRDKDQKRIQWIHNEYSALAGKNDWT----KRQITFTTPANAGK  1421
Db  1489 AEDGENALLKSTSESSRIKWGVTRSLDRE--IVRNEEGQVIKDHSRGYTDNFMGKGEQAI  1546

Qy   538 ---TLFQYNTDKSELGRL-LKQTECTKGKNGKTYSVVHKFTYTKQDDTLQQS-HSITTHD   592
Db  2274 -----------VRSPKKVIKQPNGATKYVGKKATVVLNKRGKVITAYG-------------SS  2312

Qy  1422 AVVYMEVDHKDKDGKGKAWFDEVQLEKGEVSSSTNPVQNSSFTSATENWNVSGASVDSEE  1481
Db  1547 LVHGDKDGFLYHTEGNK  1563

Qy   593 NFTIHRSQVRSRYTGRLFSDTDTKDIVTQMSYDKLGRLLTRT----------------L   635
Db  2313 RAKGSKHVFHTHGKGNK  2329

Qy  1482 GFNDDVSLKAARTSASQAGSVTKQTVVLGQSANDKPVYLTLTGMSKASSVKFTDEKDYSL  1541

Qy   636 NSGTPYANTLTYDYEL--NNLQDDNRPPVI  664
Db  1542 QANVTYADGSTGIYNAKFPSGTQEWNRAAVIPKTKPINKVDISILFQKSATGTVWFDDI  1601

Qy   665 ------TTTDVNGNQLRNEFDGAGRHVSQCLKDSDGDGKFYTIHTQQYDEQGRHHT   714
Db  1602 RLIEGSLLTKSTYDSNGNYVTKEEDELGYATS---TDYDETGK----KTSETDAKGEKTT  1654

Qy   715 STY--SDYLTNGRQQTPDPDKVHLSMSKSYDNWG-QIANT---------HWSYGVSEKI-   760
Db  1655 YTYDQADQLTNMTLSNGTSILH-----SYDKEGNEVSKTIRAGADQTYKFEYDVMGKLVK  1709

Qy   761 TVDPITLTATKQLQSNSNNVQT----GKEVTTYTPSQQPIQITLFDEAGHLQSCHTLTRD  816
Db  1710 TTDPLGNVLASEYDANSNLTKTISPNGNEV---------------------SLSYD  1744

Qy   817 GWDRVRKETDAIGQCTIYQYDNYNRVIQTLPDGTIVNRKYAPFSTDTLITDIRVNGISL  876
Db  1745 GTDRVKSKSYNGTEKYIFTYDKNGN---------ETSVVNKEQN------------TT  1781

Qy   877 GQQTFDGLSRLTQSODGGRVWAYTYSAGNDQCPSTVITPDGQFIHYQYQPELDDAVLQVA  936
Db  1782 KKRTFDNKNRLTELTDRGGSQTWTYPSDSDKLKTF------SWIH--------------G  1821

Qy   937 SNEITQQFSYNPVTGALLKAVAEGQSLTPIYYPSGRLKMENINDMKKMSYLWTLRGLENG  996
Db  1822 DQKGTNQFTTN-----------------------KLDQMIBMKDSYSSYSFDYDEN-  1854

Qy   997 YTDLTGTIQKISRDTHGRVTQIKDSSIKTLNYDDLNRHIGSQVTDLATGHMLTTTVEFD  1056
Db  1855 ------GNVQ---------TFITGNGGGTSFSYDERNLVSSLHIGDKNGGDILTESYEY-  1898

Qy  1057 GLNREIGRKLCDSSGHTLDIQQSWLKTQQLANRIVKLNGVLQRTEQYSYDSRNRLNQYKC  1116
Db  1899 --------DANGNRTTINSS-----------ASGKV---QYEYGKLNQLVK---------  1927

Qy  1117 DGAECPTDKYGHSIVTQNFTYDIYGNITACHTTFADGTED--HATFKFANPTDPCQLTEV  1174
Db  1928 -----ETHEDGTVI---EYTYDGFGNRKTV-TTIKDGSSKTVNASFNIMN----QLTKV  1973
```

```
Qy  1348 MRFHTPDSLSPFGAGGINPYSYCLGDPINRSDPSGHLSWQAWTGIGMGIAGLLLTIATGG  1407
Db     1 MRINCPDSWSPFGAGGINPYAYCDGDPINRVDPNGHLSWQAELDIGLGVGVLVAVFTAG   60

Qy  1408 MAIAAAGGTAAAIASTSTTALAFGALSVTSDITSIVSGALEDASPKASSILGWVSMGMGA  1467
Db    61 TSIAAAGAISAAIBSASAISLVVGTLGVAADVASIASAGALEDANPQASATLGWISLGLGG  120

Qy  1468 AGLAB-----SAIKGGTKLATHLGAFAEDG---------ENALLKSTSESSRIKWGVT--RS  1513
Db   121 PGAVSGLATAARAGKKL---ISGLAKGGGKIRSQSPVQGISYRSLLSRGDPLRGGPHFQS  177

Qy  1514 LDRBIVRNEB------------------------GQVIKDHSRGYTDNF-----------MGK  1541
Db   178 LSRVTVAPESMRPAGLNYWHKVSQKSSLGYQHVFGADREIFGYEIREPIEFPRRRPSITK  237

Qy  1542 GEQAIL------------------VHGDKDGFLYHTEG-----------NKHNGKGPYTRHTPEQLVDY  1581
Db   238 RDIVILSGTHGRVHGDN------WTSQLRRPDILERAFYIEDVQNYKG-----------QLNGR  285
```

```
Qy   1582 LKDNNIVDLTQGG--------DKPVHLLSCYGKSSGAADKMAKYIN-RPVIAY 1625
          :|: .:|:.           ::.:|| :|||||:         |||
Db    286 VKIVDMAGMTESEFGRHVRNVNQHVILGYCWGRNDEA---LTYHRNLRPVISY 335

RESULT 5
AF1489
cell wall-associated protein precursor wapA (B. subtilis) homolog lin0454 [imported]  - L
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C;Accession: AF1489
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AF1489
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-2167 <GLA>
A;Cross-references: UNIPROT:Q92EK5; GB:AL592022; PIDN:CAC95686.1; PID:gi16412895; GSPDB:C
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: lin0454

Query Match        4.3%;   Score 386;  DB 2;  Length 2167;
Best Local Similarity  19.4%;   Pred. No. 1.7e-12;
Matches  327;  Conservative 212;  Mismatches 582;  Indels 568;  Gaps    67;

Qy     17 NNEFFTQANNFTSAVSGGVDPRTGLY---NIQITLGHI-VGNGNL-------GP 59
             :    |.  :  ::.   ::. . : .::  :. . : ||.||| ||
Db    718 NNTTSVQSEPATPSIPERALNKQLGMVDYWTSVPVRGGEVNATNGNFLFHETDFNLEGRGP 777

Qy     60 TLPLTLSYSPLNKTDIGFGIGFNFGLSVYDRKNSLLSLSTGENYKVIETDKTV----- 112
           |.:. .:  |.       ||| ::|..      .|:|.| : ||:|:.
Db    778 SINVNRTFNSQDDATGIFGKGWTSTL----EEKLVEEENG-NIVWVESDKKIIHRFTKKG 831

Qy    113 ---------KLQQKKLD------NLRFEKDLKENCYRIIHKSGDIEV 144
                     ||.|||:       .:|::|:.|  |  :::| :||
Db    832 DKYEAPPGIYSEITKNADGYLKIEEDKSETRFLVDGRLKSEKDTKGNLTYEYTDGKLTS 891

Qy    145 LTGFNNNAFDL-----KVPKKLLNPAGHAIYIDWNFEATQPRLNRIIYDDLDLDGHDIPLLNLE 200
          : :   |.|:.    :.  .::.| | :|:|:. |.|.:   |.:|||:|:||.||.||||:||
Db    892 LRDASGRTVTLTYEGELVKELVGPEDRKISYTYN-------LTSIYDPKHTEEKPYETTFAYEE 972

Qy    201 YQGLIKTLILTLFPGQKEGYRTELRFLNRQLNSIHNFSLGNENPLTWSFGYTPIGKNGILG 260
          :  ||||.:..  .   ||.|:|:  ::|..|| .|: || |  |.::  |.  |.||:||
Db    927 KQELISSSTARGKLYRYGYTDGL----LTSIYDPKHTEEKPYETTFAYEE------SDG 1864

Qy    261 QWITSMTAPGGLKETVNYSNNNQG----HHFPQSANLPVLPYVTLMKQVPGA 308
          : || :|| ||.|:.|.|:||:.:    .:.||:|.|:|.:||:|.:||.|.
Db    973 EKLTEITDPVGKKTTLSYDKAEQTTLTNEKKKKTIYSYNDAGN--------PKKEIVDA 1024

Qy    309 GQPAIQAEYSYTSHSNYVGGSNGIWNNKLDNLYGLMTEYNYGSTESRRYKDKEGH--- 363
          |:|||:||.|.|:.||:.|| .:|:|:||:||||.|:||||:|||||.||::.:.||:
Db   1025 DGLKLTTTYTYESNN-----LVKEVNPKGQEETYAYADADGNTTKAT 1065

Qy    364 DQIVRIERTYNNVHLLTSECKQQNGYIQTTETAYYAIIGHNFDSQPSQFQLPKTKETWR 423
          |.:|  : ||:|||:.|:.:|:| ||.|:.| || .:.||||.|||.|||||:.|:||.:
Db   1066 DAYGTESYTYNDNNDVIS-----STDTE-----GRKTTTVTYD 1097

Qy    424 SADNSYRSEITE-----TTFDESGNPL-------------TRKVIK 450
          ::||:||:.||:     ::||:::||:             |:||:|
Db   1319 ILGYNGERA--------DPVSGVTHLGNGYRSYDPTLMRFHTPDSLSPFGAGGINPY 1367

Qy    451 DKK------TQ-------KIISP--STHWEYYPPAGEVDNCPPEPYGFTRFVKK 489
          |.|      ||       |::|:  |||.|||::.  ||.|.|..||.|..|.|
Db   2001 YAGYTYDKEIEQYYLMARYYEPQGV----FTAYDP----YPGDEDDP----QTMNGY 2046

Qy   1158 DAKGSMTFDNTQSAPGALGGSGSVKLTSEANSTVKGYSSVTQRVDVEPETTYTFSAWIKT 1217
          |.||:||||||| || |.|:||: :|.:|   :.:|||:|.:|...|| |:|.:||.|||
Db   2047 NYANNNPVMFDPDGNVAWWIAAAGYGAFEGGAEYLLLTTKKKNWK----GFGKAVVKGA 2101

Qy    490 --------IIQTPYDSEFKDDPEKFIQYRYSLIGSQSHVTLKIEER-HYSATQLLNSTL 539
                  :|||||.:|.:||::||:||||:|||||||||||||::: ||:|||||||||
Db   1425 TTALAFGAL 1433
```

```
Db  2102 VLGLGFGKL 2110

RESULT 6
T37218
hypothetical protein SC2H4.02 - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
A;Accession: T37218
R;Oliver, K.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, September 1998
A;Reference number: Z21615
A;Accession: T37218
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2183 <OLI>
A;Cross-references: UNIPROT:O86585; EMBL:AL031514; PIDN:CAA20596.1; GSPDB:GN00070; SCOED
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SC2H4.02

Query Match    4.3%;  Score 381; DB 2; Length 2183;
Best Local Similarity  19.8%;  Pred. No. 3.2e-12;
Matches  366; Conservative 221; Mismatches 669; Indels 594; Gaps  91;

Qy   99 TGENYKVIETDKTVKLQQKK---------LDNLRFEKDLKENCYRIIHKSGDIEVLTGF 148
         :||:  :|||  |::|||||         |:|   :    :  |:::|::|| |::|:|
Db  620 TGDS-----TDQSLWLDEIKHTGKRGTDLSLDAVKFSHVFMPN--RVDGPSDDI------ 666

Qy  149 NNNAFDLKVPKKLLNPAGHAIYIDW----------NFEATQPRLNRIYDDLDGHDIPL 196
         : :  ::  |  :::| |::|:|            :|  | || :|:|| |:|  |||
Db  667 --LSFERPRLRTVVSETGAQTIVDYLPADCVAGQTMPKADENTKRCFPVVWSPYGQEEPI 724

Qy  197 LNLBYQGLIKTILTLFP-GQKEGYRTELRFLNRQLNSIHNFSLG------NENPL----- 244
         |:  :: ||   | | | |:    :|   :  :::| :  :|       :||:|
Db  725 LDWFQKYPVSSVRTTDPLGGSEA-------VQHTYEYSGGGAWHYNDDPLTPAKER 773

Qy  245 TWSFGYTPIGKNGILGQW------ITSMTAPGG----LKETVNYSNNNQGHHFPQSANLPVL 296
         ||:             |:      ::| | ||     |||  :||:|:::|||| :||: |:
Db  774 TWS---------QWRGYGKVTHYTGPSGGTRAKTVTVYLRGMDGDRVLGGDGKTPD 820

Qy  297 PYVTLMKQVPGAGQPAI----QAEYSYTSHNYVGGGSNGIWNNKLDNLYGLMTE 346
         |  |: ::   | || :    ||||:|:||||| |||:||:|||||||||||:|
Db  821 PDRRRKAEVSGVTAAAVTDSDQLAGFQRESVAYDGDKEVSGTVNDPWSKRT--------- 871

Qy  347 YNYGSTESRRYKDKEGHGDQIVRIERTYNNYHLLTSECKQQNGY--IQTTETAY--YAIIG 402
         :|:|:|:|||| ::     |:||:  ||::|:::|:   ::||:  |:|| | :  |||:|
Db  872 ---ATQHKSYADTRAY--YVRVGAS----HARTRITSRLNPYDRVRTTRTGYDDYGMPV 921

Qy  403 HNFDSQPSQFQLPKTKTETWRSADNS----------YRSEITETTFPD---ESGNPL 445
         |:     :: ::|    | |::|| :          |||||||| |||   |||:||
Db  922 SVEDLGDDSVTGDEKCTRTWYARNDAAGLTALVSRTRIVGRACSVTDSALDLPADAARP 980

Qy  446 TKVIKDKTQKIISPSTHWE--YYPPAGEVDNCPEPYGTRFVKKIIQTPYDSEFKDDP 503
         |:|:  ||:|::||::|:|  ||| |  :   | :|  |  :|||||| |:|||||| TA
Db  981 GDVVSDKAT----AYDTTTWSATQKPTKGDA-RWSGRAKGYGGDRPLWQTTAVTDY-DTL 1035

Qy  504 EKFIQYRSLIGSQSHVTLKIEERHYSATQLLNSTLFQYNTDKSELGRILLKQTECTKGEN 563
         ||:  |:| ||:|:|::|:||||:|||||||||:||||||||||||||||||||||||:|
Db 1036 GRPVQVR----NTDDVITSKTEYQPPAAGPLTSTTVV----------N 1069

Qy  564 GKTYSVVHKFTYTKQDDTLQQSHSITTHDNFTIHRSQVRSRYTGRLFSDTDTKDIVTQMS 623
         ||:|::  :|||||||||||:|:|| ||||||||||:||||||||||:||||||||||:|
Db 1070 AK-----------GHRTTTVKDFAL----GADLKVTDANGKVTESA 1100

Qy  624 YDKLGRLLTRTL-NSGTPYANTLTYDYELNNLQDDNRPPVITTTDVNGNQLRNEFDGAG 682
         ||:|||:::| | ||| | | ::|:|:|:||||||:|||::|||||||:|||||||||:
Db 1101 YDSLGRVTSVWLPNRSRALGKTANVVYGYS--VKSTSLPWVSSAT-LNG------DGSG 1150

Qy  683 RHVSQCLKDSDGDGKFYTIHTQQYDEQGRHHTSTYSDYLTNGR---------- 725
         ||:|:  :|| || :: ||:|||||||||||:||:||||||||
Db 1151 YRTTYEIYDS------LLRTRQVQAPSAQGRVIAQTLYDGRGRLPVTAQADIWDDTAAP 1203

Qy  726 ---QQTDPDKVHLSMSKSYDNWGQIAN-THWSYGVSE----------------------------------------------------------------- 758
           :||:||||||:|:|||||||||||| :||||||:|
Db 1204 AGKIVQIDGGQAPRQTDSVVDGMGRVTKAVTKSYGVTEWAVDTAYRGDLVLTGAPEGSA 1263

Qy  759 -KITVDPITLTATKQLQSNSNNVQTGKEVTTYTPSQQPIQITLTDFEAGHLQSCHTLTRDG 817
          |:|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1264 NAVVTDAFGRTVERRDYAGTQPAGTDVMTTRYA--------------------------------------------------------------------------- 1313

Qy  818 W-----------------------DRVRKETDAIGQCTIYQDYNYNRVIQITLpD 849
         |                       :|| :||  ||| ||::|||||||||||:|
Db 1314 WTYTYDLFGRQVSVTDPDKGTTVTEYDALDRAVKSTDGRGEVLLFEYDVLGRKTGMWQSA 1373

Qy  850 GTIVNRKYAPFSTDTLI-------TDIRVNGISLGQ-------QTFDGLSRLTQSQ----- 891
         ||:|:||||||||||||       |||||||||||||       |||||||||||||
Db 1374 KTGAN-KLAAWSFDTLAKGQQDTAVRYEGGETGRAYTQKVTRYDPLYKVTNNELTLPAND 1432

Qy  892 ---DGGRVWAYTYSAG-------NDQCPSTVITPDGQFIHYQYQPELDDAVLQVASNEIT 941
         :::|||||||||||||       |:|||||||:||||||||||:|||||||||||||||
Db 1433 PLVAAGVPARLAFSTGYNLDGTVKQAAAPAVAGLSAETVSYTY------DGLGQVLTAKGT 1487

Qy  942 ---QQFSYNPVTGALLKAVAEGQSLTPIYYPSGRLKMENINDMKKMSYLWTLRGLGENGY 997
         :::||||||||||||||||||||||||:||||||:::::::::::::::||||||||||:
Db 1488 TGYLQQAYSP-----------LGDLRQWTLATDFTGAKKVYLNND------YEAGTRRLTRSY 1534

Qy  998 T--DLTGTI----QKISRDTHGRVTQIKDSSIKTTLNYDDLNRHIGSQVTDLATGHMLTTT 1052
         |  |:|||:    |||||||||||||||||||||||||||||||||||||||||||||||
Db 1535 VTDDVHGFMLQELKYQQDDAGNITSVSDA---TTLG--------------GTGKADHQC 1576

Qy 1053 VEFDGLNREIGRKL-------CDSSGHTLDIQQSWLKTQQLANRIVKLNGVLQRTEQYSYD 1106
         ||:||||||||:||       |:|||:||||||||||:||||:|||||||||||||||:|
Db 1577 FTYDG-HRRLSEAWTPETADCSTSGRT-----------------------VAGLGGAAPYWTSYQYD 1619

Qy 1107 S---RNRLNQYKCDGAECPTD-KYG-------HSI----VT----QNFTYDIYGNITACHTT 1149
         :   |||||||||||||||| |||       |:|    |:    ||:|||:||||||||||
Db 1620 DSGLRSKQTEHRMSGDDVTTBYEYGTAEGQPHALSATVTGAENASYTYDETGN-------- 1672

Qy 1150 FADGTEDHATFKFANPTDPCQLTEVHHTHPDMPDNIRLKYDKAGRVINITDNHGNTENFT 1209
         |||||:|||||||:||:||||||||||:|||||||:|||||||||||:::|||:|||||
Db 1673 -----------------------TE----TRPGVRATQTLDWNAEGRLAGVSEPAAGGKPAT 1707

Qy 1210 YDTLGRLQNCQGSVGYDPLNRLVSQKTDTLDCE--LYYRETMLVNEVR-NGEM-------I 1261
         |||||||||||||:|||||||||||||||||||  ||||||||:|:|: |||
Db 1708 --------------------GTAYVYDAGGDLLIRRPTTTDGETVLYLGTTEVHLKVSGNGAAKALSGA 1756

Qy 1262 RLLRTGETIIAQQRASKVLLTGTDSQQSVILTSDKQNLSQEAYSA---------------YGK 1309
         ||||||||||||||||||||||||||||||||||||||||||||:               ||
Db 1757 RTYKAGSAVIAVRTST---AGVSGYKLTFLAGDHHGTSGLAINADTLAFAKRWSTPFGA 1812

Qy 1310 HKSTANDA----SILGYNGERADPVSGVTHLGNGYRSYDPTLMRFHTPDS-LSPFGAGGINP 1366
         |:||||||    |:||::|:|||::|:|||||||||||||:||||||||| |:||||||:||
Db 1813 PRGTASGAWPDDRGFLGKPADAATGLTQL--GARQYDPDTGRFLSVDPLLEPDKPNTLNG 1870

Qy 1367 YSYCLGDPINRSDPSGHLSWQAWTGIGMGIAGLLLTIAT--GGMATAAAGGIAAAIASTS 1424
         |:|:  ||:|||||||||||:::||:|:|||||||||:||  ||||| |||:|||:: TA
Db 1871 YAYASNSPVTNSDPSG---------TSDGLG--GLLGAIGAIIGGVVGAVIGAVGAAI--TA 1919

Qy 1425 TTALAFGALSVTSDITSIVSGALEDASPKASSILGWVSMGMGAAGLAAESAIKGGTKLATH 1484
         |||||||||||||||||||||||||||||||||||:||:|:|:|:|:||||||||||:|:|
Db 1920 ----------------GGWGGTAPTSSG--GW-----------------------TQPLTKQWTPGATY 1956

Qy 1485 LGAFAEDGENALLKS----------TSES----------SRIKWGVTRSLDREIVRNEE--------- 1523
         |:|: |:|:|||||:          ||||          |||||:|||||||||||||
Db 1957 NFITKSWDLPFNPPSQSLEEMLASMPDWGIVS--DPKAANRWETSRSLFFG 2005

Qy 1524 GKTYSVVHKFTYTKQDDTLQQSHSITTHDNFTIHRSQVRSRYTGRLFSDTDTKDIVTQMS 1574
         ||:  |:   :||::|| :||::  |::|:|::::|:  |::|:|| :|::|:::||:|:|
Db 2006 WLWGGGYPLRRHQDFRG------GDAFTSIIAQDETISGLRSKMVCQARKGTKGAFAKEV 2059

Qy 1575 PEQLVDYLKDNNIVDLTQGGDKPVHLLSCVGKSSGAADKMAKYINRPVI-------- 1623
         |||||||||||||||||||||||||||||||||||:||||||||||||
Db 2060 GFQYVD-----------EGPEPGSPWYKFNSLRGAAKDIAGVLTNGAVGTGNQADAFLG 2107

Qy 1624 AYSNKPTISQGLARIERKDFFLKSTYHSYDPRKIILGRTEKTVKPKTFRP 1673
```

```
Db   1038 ----------------SQVTDLATGHMLTTTVEFDGLNREIGRKLCDSSGHTLDI---QQS 1079
                                                           :::|:::|
Qy    721 PAEQWQYDEHGWLTEISHLSEGHQVAVHGYDYDKGRLAGERQTVHNPETGELLWQHETEH 780

Qy   1080 WLKTQQLANRIVKLNGVLQRTEQYSYDS-------------RNRL------N 1112
             :| |:   |:  :|:|:| :| ||                |||:      |
Db    781 AYNEQGLANRVTP--DSLPRVEWLTYGSGYLAGMKLGGTPLVEFTRDRLHRETVRSFGNN 838

Qy   1113 QYKCDGAECPTDKYGH---------------SIV----TQNFTYDIYGNIT 1144
               |:  | :   :|                 |:|    ||:| ||:||:|
Db    839 AYELTSTYTPA---GHLQSQRLNSQVYDRDYDWNDNGDLVRISGPRQTWEYGYSATGRLE 895

Qy   1145 ACHTTFADGTEDHATFKFANPTDPC--QLTEVHHTHPD----MPDNIRL---KY 1189
               |  | |:|   |  || | |:  | |||:|  ||      |||::|    ||
Db    896 SVRTLASD-----LDIRIPYATDPAGNRLPD-PELHPDSTLTAWPDN-RIAEDAHYVRH 948

Qy   1190 DKAGRVINITD------NHGNTENFTYDTLGRL------ONGQGSV---YGYDPL-- 1229
              |  | |  |       :|| :|::|  ||        :    |    | ||
Db    949 DEYGRLTEKTDRIPAGVIRTDDERTHHYHYDSQHRLVFYTRIQHGEPLVESRYLYDPLGR 1008

Qy   1230 ------------------NRLVSQKTDTLDCELYR------- 1247
                                :||:  :|:|  |:|
Db   1009 RMAKRVWRRERDLTGWMSLSRKPEVTWYGWDGDRLTTVQTDTTRIQTVVEPGSFTPLIRV 1068

Qy   1248 --------ETMLVNEVRNG---------EMIRLLRTGETIIAQQRAS---------- 1277
                      |::|:||::        :| :|||  ||:|
Db   1069 ETENGEREKAQRRSLAEETLQQEGSENGHGVVFPAELVRLLDRLEEEIRADRVSSESRAWL 1128

Qy   1278 ------------------KVLLTGTDSQ--QSVILTSDKQNLSQEAYSAYGKHKST 1313
                                |:| | |   :  |  ::  ||::|
Db   1129 AQCGLTVEQLARQVBPEYTPARKVHFVHCDHRGLPLALISEDGNTAWRGEYDEWGNQLNE 1188

Qy   1314 ANDASI--LGYNGERADPVSGVTHLGNGYRSYDPTLMRFHTPDSLSPPG-AGGINPYSY 1369
              |  |:  |   |  | | : | |   |  | | | |  | ||||  |  | :||:||
Db   1189 ENPYLHQPYRLPGQQHDEESGLYYNRNRY---YDPLQGRYITQD---PIGLAGGWNLYNY 1243

Qy   1370 CLGDPINRSDPSG-------HLSWQAWTGIGMGIAGLLLT-----IATG--GMAIAAAGGI 1416
                :|| |    ||        :|  | |:|  |:   |:|       ||  || |  | |
Db   1244 PL-NPIIRMDPLGLYNLYQLLYDVWHDDSYGTSSIDITGSGDLISLGGHAGLGVAFAKKK 1302

Qy   1417 AAAIASTSTTALAFGALSVTISDTTSIVSGALEDASPKA---SSIIGWVSMGMGAAGLAESA 1474
              ||  | :|:  | |                  |   ||    :: |  | :| |    :|
Db   1303 GEMLSDICIYATACGHAGIGGGINAAITYSYKSLPTSGVSNVG-VTVGGGVGG------ 1356

Qy   1475 IKGGTKLATHLG-AFAEDGENALLKSTSESSRIKWGVTRSL 1514
              :  | | ||       : |:  || ::::| :||||  :|
Db   1357 -------HFAYTYVVVDVDNP--ESSTESVGIGAGVDASV 1386
```

RESULT 8
AB0539
Rhs-family protein [imported] - Salmonella enterica subsp. enterica serovar Typhi (strai
C;Species: Salmonella enterica subsp. enterica serovar Typhi
A;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AB0539
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov.
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AB0539
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-843 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD08754.1; PID:g16501575; GSPDB:GN00176
C;Genetics:
A;Gene: STY0324

Query Match   4.0%; Score 353.5; DB 2; Length 843;
Best Local Similarity   22.8%; Pred. No. 2.1e-11;

```
Db   2108 TYSGKARISS-----INKKEGSVTLKFSAWNGSD---WRSATHVVPRSWNP 2150
```

RESULT 7
E90886
RhsE core protein with extension (imported) - Escherichia coli (strain O157:H7, substrai
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C;Accession: E90886
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: E90886
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1400 <HAY>
A;Cross-references: UNIPROT:Q82X2F8; GB:BA000007; PIDN:BAB35484.1; PID:g13361527; GSPDB:G
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs2061
C;Superfamily: rhsF protein

Query Match   4.1%; Score 365; DB 2; Length 1400;
Best Local Similarity   20.9%; Pred. No. 1.1e-11;
Matches 284; Conservative 154; Mismatches 467; Indels 456; Gaps 65;

Qy   418 KTETW-RSADNSVRSEITETTFPESGNPLTKVI------------KDKKTQKIISPSTHWEY 466
              :| |::|| :| |:|: | ||: |:| |||| | ||:| :|
Db   218 RTLTYRRAAGDLAGEITGVT-DGAGREFRLVLTTQACRAEEARKQHTASLSSPDT---- 272

Qy   467 YPPAGEVDNCPPEPYGFTRFVKKIIQTPYDSEFKDDPBKFIQYRYSLIGSQSHVTLKIEE 526
                                    |::|:| || :|:
Db   273 -----------PRPLSDSAFPDTLPGTEY-------GPDRGIR------ 297

Qy   527 RHYSATQLLNSTLFQYNTDKSELGRLLKQTECTKGENGKTYSV-----VHKFTYTKQD 579
              :|| |::|: ||: | | |||| | : | || ||
Db   298 --LSAVWLTHDPAYPESLLPGAPLARY------TYTEAGBLLAVYDRSNTQVRAFTYDAQH 349

Qy   580 DTLQQSHSITTHDNFTIHRSQVRSRYTGRLFSDTDTKDIVTQMSYDKLGRLLTRTLNSGT 639
              | | ||| || :| | | | |:|:| : | | :| ||| ||
Db   350 PGRMVAH------------------RYACR---------PEMRYRYDDTGRVVEQLNPAGL 383

Qy   640 PYANTLTYDYELNNLQDDNRPPFVITTTD-VNGNQLRNEFDGAGRHVSQCLKDSDGDGKF 698
              : | : ||:: |::| | || | ||: |::| || ||:|
Db   384 SY-----RYQYE-----QDR------ITVTDSLNRREVLHTEGGAG--LKRVVKKELADG-- 425

Qy   699 YTIHTQQYDEQGRHHTSTYSDYILTNGR---QQTDPDKVHLSMSKSYDNWGQIANTHWSYGV 756
                |:| |:: | |:|| || ||::| | | | | : | | |
Db   426 --------------SVTHSGYDAAGRLLTAQTD------------------AAGRRTEYGL 453

Qy   757 SEKITVDPITLTATKQLQSNSNNVQTGKEVTVTPSQQPIQITLFDEAGHLQS---CHT 812
               ||:| || |:| : | :: | :| :| ||:|| || ||
Db   454 N---VVSGDITDITTPDGRETKFYNDGNQLTAVVSPDGLESRRAYDEPGRLVSETSRCGD 511

Qy   813 LTRDGWDRVRKE------TDAIGQCTIYQDNYNRVIQITLPDGTIVNRKYAPFSTDTLIT 867
              | ||| | | | : | || :| | | ||||: |: | |||
Db   512 VIRYAYDNPHSELPATTTDATGSTRQMTWSRYGQLLAFTDCSGYQTRYEYDRFGQMTAVH 571

Qy   868 DIRVNGISLGQQOTFDGLSRLTQSODG-GRVWAYTYSAGNDQCPSTVITPDGQFIHYQYQP 926
              |: | || | | || ||: |: | | :| | :| : | || | || | | ||
Db   572 --REEGIS-RYRRYDNRGRLTSVKDAQGHETRYEYNAAGDL--TAVITPDGNRSETQY-- 624

Qy   927 ELDDA---VLQVASNEITQQFSYNPVTGALLKAVAEGQSLTPIYYPSGRLKMENINDMKK 983
              || | |: | |::| | | | | :::| ||| | | | ||
Db   625 ---DAWGKAVSTTQQGLTRSMEYD-LAGRITTLTNENGSRSEFTYDA------------- 667

Qy   984 MSYIWTLRGLENGYTDLTGTIQKISRDTHGRVTQIKDSSIKTTLNYDDLNRHIG------- 1037
              || | | | | |||:| | :| |:||:| | | :| | | | ||||::|
Db   668 LDRLVQQRGFD-------GRTQRYHYDLTGKLTQSEDEGLVTLMHYDESDRLLTHRTVNGE 720
```

rhsC protein precursor [similarity] - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence revision 17-Sep-1997 #text_change 09-Jul-2004
C;Accession: C64805; I69400; T48912
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C
.A.; Rose, D.J.; Mau, B.; Shao, Y.
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: C64805
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1397 <BLAT>
A;Cross-references: UNIPROT:P16918; GB:AE000173; NID:g1786910; PIDN:AAC73794
A;Experimental source: strain K-12, substrain MG1655
R;Feulner, G.; Gray, J.A.; Kirschmann, J.A.; Lehner, A.F.; Sadosky, A.B.; Vlazny, D.A.;
J. Bacteriol. 172, 446-456, 1990
A;Title: Structure of the rhsA locus from Escherichia coli K-12 and comparison of rhsA
A;Reference number: I54935; MUID:90094253; PMID:2403547
A;Accession: I69400
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1221-1397 <RES>
A;Cross-references: GB:M29718
A;Note: the sequence is revised in GenBank entry ECORHSCA, release 114. (PIDN:AAC63073.
T;Oshima, T.; Aiba, H.; Baba, T.; Fujita, K.; Hayashi, K.; Honjo, A.; Ikemoto, K.; Inad
T.; Mizobuchi, K.; Mori, H.; Motomura, K.; Nakamura, Y.; Nashimoto, H.; Nishio, Y.; Sa
DNA Res. 3, 137-155, 1996
A;Title: A 718-kb DNA Sequence of Escherichia coli K-12 Genome Corresponding to the 12.
A;Reference number: Z25006; MUID:97061202; PMID:8905232
A;Accession: T48912
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1397 <OSH>
A;Cross-references: EMBL:D90709; NID:g1651305; PIDN:BAA35359.1; PID:g1651309
A;Experimental source: strain K12; Kohara clone 174
C;Genetics:
C;Gene: rhsC
A;Map position: 15.7-16.0
C;Superfamily: rhsF protein

Query Match          3.9%;    Score 342;    Pred. No. 2.1e-10;    Length 1397;
Best Local Similarity   21.4%;
Matches   256;    Conservative 157;    Mismatches   382;    Indels   404;    Gaps   63;

Qy    418    KTETW-RSADNSYRSEITETTFDESGNPLTKVI---------PS------   461
             :::|:|::|:| ::::::|::| |::||:::
Db    218    RTQTFHREAAGEFSGEITGVT-DGAGRHFRLVLTTQAQRAEARQQAISGGTEPSAFPDT   276

Qy    462    --------------------------THWEYYPPAGEVDNCPPEP----YGFT------   488
                                       |:|:|:
Db    277    LPGYTEYGRDNGIRLSAVWLTHDPEYP-----ENLPAAPLVRYGWTPRGELAAVYDRSNT   331

Qy    489    KIIQTPYDSEFK------------DDPEKFIQYRYSLIGSQSHVTLKIEERHYSATQLLNSTL   539
             :: |||
Db    332    QVRSFTYDDKYRGRMVAHRHTGRPE--ICYRYD---SDGRVTEQLNPAGLSYT-------   379

Qy    540    FQYNTDKSELGRLLKQTEC--TKGENGKTYSVVHKFTYTKQDDTLQQSHSITTHDNFTIH   597
             :||| ||:| :|:
Db    380    YQYEKDRITITDSLNRREVLHTQGEGG-LKRVVKK----------EHADGSVT   421

Qy    598    RSQVRSRYTGRLFSDTDTKDIVTQMSYDKLGRLLTRTLNSGTPYANTLTYDYELNNLQDD   657
             :||::|::| ::|:::|:: ||
Db    422    QSQFDA--VGRLRAQTDAAGRTTEYSPDVVTGLITRIT--TPDGRASAFYY-------   468

Qy    658    NRPPFVITTTDVNGNQLRNEFDGAGRHVSQCLKDSD-------GDGKFYTIHTQQYDEQGRH   712
             .||:|::|||:
Db    469    NHHSQLTSATGPDGLEIRREYDEWGRLIQETAPDGDITRYRYDNPHSDLPCATEDATGSR   528

Qy    713    HTSTYSDYLTNGRQQTDPDKVHLSMSKSYDNWGQIANTHWSYGVSEKITVDPI-TLTATK   771
             |:|:|||||||||| ::|:
Db    529    KTMTWSRY---GQLLSFTDCSGYVTRYDHDRFGQVTAVHREEGLSQYRAYDSRGQLIAVK   585

RESULT 9
C64805

```
Query Match          3.8%;    Score 341.5;   DB 2;   Length 1404;
Best Local Similarity  20.2%;   Pred. No. 2.3e-10;
Matches 255;   Conservative 130;   Mismatches 370;   Indels 505;   Gaps 59;

Qy  376  YHL-LTSECKQQNGYIQTETAYYAIIGHNFDSQPSQFQLPKT-KTETWRSADNSYRSEI  433
Db  245  FHLVLTTQAQRAEVFRKQRATSLSSPAGPR--SASSSLVFPDTLPAGTEYGADNGIRLEA  302

Qy  434  TETTFDES---------GNPLTKVIKDKKTQKIISPSTHWEYYPPAGEV----DNCPPEPYGF  483
Db  303  VWLTHDPAYPDELPAAPLARYT----------------YTASGELRAVYDRSGTQVRGF  345

Qy  484  TRFVKKIQTPYDSE----------FKDDPEKFIQYRYSLIGSQSHVTLKIEERHYSATQL  534
Db  346  A------------YDAEHAGRMVAHHYAGRPES--RYRYDDTG--------RVTEL  379

Qy  535  LN----STLFQYNTDKSELGRLLKQTEC-TKGENGKNTYSVVHKFTYTKQDDTLQQSHSI  588
Db  380  VNPEGLDYRFEYGQDRVTITDSLNRREVLYTEGEGG-LKRVVKK----------  422

Qy  589  TTHDNFTIHRSQVRSRYTGRLFSDTDTKDIVTQMSYDKLGRLLITRTLNSGTPYANTLTYD  648
Db  423  -EHADGSITRSE-----------------YDEAGRLKAQT-------------  444

Qy  649  YELNNLQDDNRPPFVITTDVNGNQLRNEFDGAGRHVSQCLKDSDG----------DGKFYT  700
Db  445  --------------------DAAGRRTEYSLHMASGAVTAVTGPDGR--T  472

Qy  701  IHTQQYDEQGRHHTSTYSDYLTNGRQQTDPDKVHLSMSKSYDNWGQIANTHWSYGVSEKI  760
Db  473  VR-YGYNSQRQVTSVTYPDGLRSSREYDEKGRLAAETSRSGE-----TTRYSYD------  520

Qy  761  TVDPITLIATKQLQSNSNNVQTGKEVTYTPSQQPIQITLFDEAGHLQSCHTLTRDGWDR  820
Db  521  --DP------------ASELPTGIQDATGSTKQM--------------------AWSR  544

Qy  821  ---VRKETDAIGQCTIYQYDNYNRVIQITLPDGTIVNRKYAPFSTDTLITDIRVNGISLG  877
Db  545  YGQLLFTFDCSGYTTRYEVDRYGQOIAVHREEGISTYSSYNP----------------RG  588

Qy  878  QQTFDGLSRLTQSQDGGRVWAYTYSAGNDQCPSTVITPDGQFIHYQYQPELDDA-----VLQ  934
Db  589  Q------LVSQKDAQGRETRYEYSAAGDL-TAIVAPDGSRSEIQY--------DAWGKAVS  634

Qy  935  VASNEITQQPSYNPVTGALLKAVAEGQSLTPIYYPSGRLKMENINDMKKMSYLWTLRGLE  994
Db  635  TTQQGLTRSMGYDAAGRITVLTNENGSQSTFRYDPVDVRLTEQRGFDGRTQRY-------  686

Qy  995  NGYTDLTGTI--------------------------------QKISRDTHGRVTQIKDSS----  1022
Db  687  --HYDLTGKLTQSEDEGLVTLWHYDASDRITHRTVNGDPAEQWQVDEHGWLTTLSHTSEG  744

Qy 1023  --IKTTLNYDDLNRHIGSQ--VTDLATGHML  1049
Db  745  HRVSVHYGYDDKGRLTGERQTVENPETGEMLWEHETGHAYSEQGLATRQEPDGLPPVEWL  804

Qy 1050  ----------TTTVEF--DGLNREIGRKLCDSSGHTLDIQQSWLKTQQLANRIVKL  1093
Db  805  TYGSGYLAGMKLLGGTPLVEYMRDRLHRETARSF--GGEAYELATAWNTSGQLRSRHLNL  861

Qy 1094  NGVLQRTEQSYDSRNRLNQYKCDGAECPTDKYGHSIVTQNFTYDIYGNITACHTTPADG  1153
Db  862  P----QLDRDYDWNDNGQL--IRISGPQ-------------ESREYRYSDTGRLTGVHTTAANL  906

Qy 1154  TED--HATFKFANP-TDPCQLTEVHHTHPD----------MPDNIRL-------KYDKAGRVIN  1197
Db  907  DIDIPYATDPAGNRLPDP------ELHPDSTLTAWPDN-RIAEDAHYVYRYDEYGRLAE  958

Qy 1198  ITD------NHGNTENFTYDTLGRL------------QNGQGSV---YGYDPL--------  1229
Db  959  KTDRIPEGVIRMHDERTHYHYDSQHRLVFHTRIQHGEPQVESRYLYDPLGRRTGKRVWR  1018


Qy  772  QLQSNSNNVQTGKEVTYTPSQQPIQITLFDEAGHLQSCHTLTRDGWDRVRKETDAIGQ-  830
Db  586  DTQGHETRYE-----------------YNAAGDLTT--VIAPDG-SRNGTQYDAWGKA  623

Qy  831  -CTI------YQYDNYNRVIQTLPDGTIVNRKYAPFSTDTLITDIRVNGISLGQQTF-  881
Db  624  ICTTQGGLTRSMEYDAAGRVIRLTSENGS-----HTTFRYDVLDRLIQETGFDGRTQRYH  678

Qy  882  -DGLSRLTQSQDGQGRVWAYTYSAGNDQCPSTVITPDGQFIHYQYQPELDDAVLQVASNEI  940
Db  679  HDLTGKLIRSEDEGLV-----------THWHYD-EADRLHRTVNGET  714

Qy  941  TQQFSYNPVTGAL--LKAVAEGQSLTPIY-YPS-GRLKMENI-----NDM--KKMSY  986
Db  715  AERWQYDE-RGWLTDISHISEGHRVTVHYGYDSKGRLASEHLTVHHPQTNELLWQHETRH  773

Qy  987  LWTLRGLENGYTDLTGTIQKISRDTHGRVTQIKDSSIKTLNYDDLNRHIGSQVTDLATG  1046
Db  774  AYNAQGLAN--RCIPDSLPAVEWLITYG----------SGWLSGMKLG  808

Qy 1047  HMLTTTVEF--DGLNREIGRKLCDSSGHTLDIQQSWLKTQQLANRIVKLNGVLQRTEQYS  1104
Db  809  D--TPLVEYTRDRLHRETLR------SFGRYELITAYTPAGQLQSQ--HLNSLLS-DRDYT  858

Qy 1105  YDSRNRLNQYKCDGAECPTDKYGHSIVTQNFTYDIYGNITACHTTFAD------GTEDHA  1158
Db  859  WNDNGELIR------ISSPRQ-------TRSYSYSTTGRLTGVHTTAANLDIRIPYTTDPA  906

Qy 1159  TFKFANPTDPCQLTEVHHTHPD-------MPDN----IRLKYDKAGRVINTD------  1200
Db  907  GNRLPDP---------ELHPDSALSMWPDNRIARDAHYLYRYDRHGRLTEKTDLIPEGV  956

Qy 1201  ---NHGNTENFTYDTLGRLQNGQGSVYG--------YDPLNRLVSQKTDTLDCEL---  1244
Db  957  IRTDDERTHRYHYDSQHRLVHYTRTQYAEPLVESRYLYDPLGRRVAKRVWRERDLTGWM  1016

Qy 1245  YY------RETMLVNE------VRNGEMIRLLR-----TGETIIAQQRA-  1276
Db 1017  SLSRKPQVTWYGWGDGDRLTTIQNDRTRIQTIYQPGSFTPLIRVETATGELAKTQRRSLAD  1076

Qy 1277  ----------SKVLLTGTDSQQSVILTSDKQNLSQEA------  1303
Db 1077  TLQQSGGEDGGSVVFPPVLVQMLDRLESEILA---DRVSEESRRWLASCGLITVAQMQSQM  1133

Qy 1304  ------------YSAYGKHKSTANDAS------ILGYNG  1324
Db 1134  DPVYTPARKIHLYHCDHRGLPLALISTEGTTAWYAEYDEWGNLLNEENPHQLQQLIRLPG  1193

Qy 1325  ERADPVSGVTHLGNGYRSYDPTLMRFHTPDSLSPFG-AGGINPYSYCLGDPINRSDPSG  1382
Db 1194  QQYDEESGLYY--NRHRYDPLQGRYITQD----PIGLKGGWNFYQYPL-NPISNIDPLG  1246
```

RESULT 10
E90658

RhsG core protein with extension [imported] - Escherichia coli (strain O157:H7, substrai
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C;Accession: E90658
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
DNA Res. 8, 11-22, 2001
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: E90658
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1404 <HAY>
A;Cross-references: UNIPROT:Q8XED9; GB:BA000007; PIDN:BAB33660.1; PID:g13359693; GSPDB:C
C;Genetics:
A;Gene: ECs0237
C;Superfamily: rhsF protein

```
Qy 1230 -----------------------------NRLVSQKTDTLDCELYYR----ETMLVNEVRNGEMI 1261
                                     :||:  :  :  :
Db 1019 RERDLTGWMSLSRKPEETWYGWDGDRLTTVQTQQTRIQTVYQPGSFTPLLRRIETENGEQA 1078

Qy 1262 RLLR------TGETIIAQ-----QRASKVLLTGTDSQQS-----VILTSDKQNLS 1300
        :                                   ||||||||||||     :
Db 1079 KARHRSLAEVLQEDTGVTLPAELAVMLGRLERELRQGSVSEESQQWLAQCGLTAEQMGAQ 1138

Qy 1301 QEA-------------------------YSAYGKHKSTANDASI---LGYN 1323
        ||                                 :
Db 1139 LEAGYIPERKLHLYHCDQRGLPLGLISPGRETALTAEYDEWGNLLSETSAQPLQQSLRRP 1198

Qy 1324 GERADPVSGVTHLGNGYRSYDPTLMRFHTPDSLSPFG-AGGINPYSYCLGDPINRSDPSG 1382
        ||   :|||                                 ||||||
Db 1199 GQQYDEESGLYYNRNRY--YDPLQGRYITQD---PIGLEGGWNLYQYPL-NPIEHIDPLG 1252


RESULT 11
H91236
RhsB core protein with extension [imported] - Escherichia coli (strain O157:H7, substrai
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C;Accession: H91236
R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: H91236
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1394 <HAY>
A;Cross-references: UNIPROT:Q8X385; GB:BA000007; PIDN:BAB38287.1; PID:g13364340; GSPDB:G
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs4864
C;Superfamily: rhsF protein

Query Match             3.8%;     Score 341;  DB 2;   Length 1394;
Best Local Similarity  21.5%;    Pred. No. 2.4e-10;
Matches  156;  Conservative  260;  Mismatches  404;  Indels  392;  Gaps  63;

Qy 418 KTETW-RSADNSYRSBEITETTFDESGNPLTKVI-----KDKKTQKIIS-----PS------ 461
       :|:|:        |||:|||||                        :
Db 218 RTQYFHREEAAGEFSGEITGVT-DGAGRHFRLVLTTQAQRAEEARQQAISGGTEPSAFPDT 276

Qy 462 ------------THWEYYPPAGEVDNCPPEP---YGFT------RFVK 488
                   |||                          |:|
Db 277 LPGTEYGRDNGIRLSAVWLTHDPEYP-----ENLPAAPLVRYGWTPRGELAVVYDRSGK 331

Qy 489 KIIQTPYDSEFK---------DDPEKFIQYRYSLIGSQSHVTLKIEERHYSATQLLNSTL 539
       ||  ::                |||        :|
Db 332 QVRSFTYDDKYRGRMVAHRHTGRPE--IRYRYD---SDGRVTEQLNPAGLSYT------- 379

Qy 540 FQYNTDKSELGRLLKQTEC--TKGENGKTYSVVHKFTYTKQDDTLQQSHSITTHDNFTIH 597
       :|:            :||||||                           ::
Db 380 YQYEKDRITITDSLNRREVLHTQGEGG-LKRVVKK--------------EHADGSVT 421

Qy 598 RSQVRSRYTGRLFSDTDTDKDIVTQMSYDKLGRLLTRTLNSGTPYANTLTYDYELNNLQDD 657
       :||          :|||      ||||||                         :
Db 422 QSQFDA--VGRLRAQTDAAGRTTEYSPDVVTGLITRIT---TPDGRASAFYY-------- 468

Qy 658 NRPPFVITTTDVNGNQLRNEFDGAGRHVSQCLKDSD-----GDGKFYTIHTQQYDEQGRH 712
       |  ::  :|   ||:|||:||||||                     |:
Db 469 NHHSQLTSATGPDGLEIRREYEDEWGRLIQETAPPDGDITRYRYDNPHSDLPCATEDATGSR 528

Qy 713 HTSTYSDYLTNGRQQTDPDKVHLSMSKSYDNWGQIANTHWSYGVSEKITVDPI-TLTATK 771
       :    :|  :||           |||                    :|  :|::|
Db 529 KTWTWSRY--GQLLSFTDCSGYVTRYDHDRFGQMTAVHREEGLSQYRAYDSRGQLIAVK 585

Qy 772 QLQSNSNNVQTGKEVTTYTPSQQPIQITLFDEAGHLQSCHTLTRDGWDRVRKETDAIGQ- 830
       :|:|                                    |  |      :||:
Db 586 DTQGHETRYE-------------YNAAGDLTT--VIAPDG-SRNGTQYDAWGKA 623
```

```
Qy  831 -CTI-----YQYDNYNRVIQITLPDGTIVNRKYAPFSTDTLITDIRVNGISLGQQTF- 881
        :|:       :|||||||::                        ||:::
Db  624 ICTTQGGLTRSMEYDAAGRVIRLTSENGS----HTTFRYDVLDRLIQETGFDGRTQRYH 678

Qy  882 -DGLSRLTQSQDGGRVWAYTYSAGNDQCPSTVITPDGQFIHYQYQPELDDAVLQVASNEI 940
                                                   |||
Db  679 HDLTGKLIRSEDEGLV--------------------------THWHYD-EADRLTHRTVKGET 714

Qy  941 TQQFSYNPVTGAL-LKAVABGQSLTPIYY---PSGRL-KMENINDMKKMSYLW------ 988
        |                      ::                  |||
Db  715 AERWQYDE-RGWLTDISHISSEGHRVT-VHYGYDEKGRLTGERQTVHHPQTEALLWQHETR 772

Qy  989 ---TLRGLENGYTDLTGTIQKISRDTHGRVTQIKDSSIKTTLNYDDLNRHIGSQVTDLAT 1045
           :||                                         :|
Db  773 HAYNAQGLAN--RCIPDSLPAVEWLTYG-------------------------SGWLAGMKL 807

Qy 1046 GHMLTTTVEF-DGLNREIGRKLCDSSGHTLDIQQSMLKTQQLANRIVKLNGVLQRTEQY 1103
        |  ::                                        :   :  :|:|:|
Db  808 GD--TPLVDFTRDRLHKRKTLRRFGRYELTTAVTPAGQLQSQHLNS---------LQYDRDY 857

Qy 1104 SYDSRNRLNQYKCDGAEBCPTDKYGHSIVTQNFTYDIYGNTACHTTPADGTEDHATFKFA 1163
        :  :                                      :|||||:
Db  858 TWNDNGELIR-----ISSPRQ----------TRSYSYSDSGRLTGVHTTAA--------NLDIRIP 900

Qy 1164 NPTDPC--QLTEVHHTHPD---------MPDN------IRLKYDKAGRVINITD------- 1200
        | |                                  |||
Db  901 YATDPAGNRLPD-PELHPDSTLSMWPDNRIARDAHYLYRYDRHGRLTEKTDLIPEGVIRT 959

Qy 1201 NHGNTENFTYDTLGRLQNGQ----------GSVYGYDPLNRLVSQKTDTLDCEL------- 1244
                 :|||                  |||   ||||
Db  960 DDERTHRYHYDSDQHRLVHYTRTQYEEPLVSSRYLYDPLGRRVAKRVWRRERDLTGWMSLS 1019

Qy 1245 -----YY------RETMLVNE------VRNGEMIRLLR------TGETIIAQQRA------ 1276
             ::      |||                           |||
Db 1020 RKPQVTWYGWDGDRLTTIQNDRTRIQTIYQPGSFTPLIRVETATGELAKTQRRSLADALQ 1079

Qy 1277 --------SKVLLTGTDSQQSVILTS------------------------------ 1294
                 :|:||                                         
Db 1080 QSGGEDGGSVVFPPVLVQMLDRLESEILADRVSEESRRWLASCGLTVAQMSQMDPVYTP 1139

Qy 1295 ------DKQNL----SQEA------YSAYGKHKSTAND--ASILGNGERADPV 1330
              ||                 ||||                    |||
Db 1140 ARKIHLYHCDHRGLPLALISKEGATEWKCAEYDEWGNLLNEENPHQLQQLIRIPGQQYDEE 1199

Qy 1331 SGVTHLGNGYRSYDPTLMRFHTPDSLSPFG-AGGINPYSYCLGDPINRSDPSGHLSWQAW 1389
        ||   :                      |||                         :
Db 1200 SGLYY--NRHYYDPLQGRYITQD---PIGLKGGWNLYTYPL-SPVNGMDPLGLYEFKSK 1253

Qy 1390 TGIGMGIAGLLL 1401
        :|   |:  |:
Db 1254 NIDDIGIFALAM 1265


RESULT 12
E65145
rhsB protein precursor - Escherichia coli (strain K-12)
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence revision 17-Sep-1997 #text_change 09-Jul-2004
C;Accession: E65145; S47701; B36902; A30092; I54935
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C
 .A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: E65145
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1411 <BLAT>
A;Cross-references: UNIPROT:P16917; UNIPROT:P16918; GB:AE000424; GB:U00096; NID:g2367230
R;Plunkett, G.
submitted to the EMBL Data Library, March 1994
```

A;Reference number: S47666
A;Accession: S47701
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1129,'Q',1131-1411 <PLU>
A;Cross-references: EMBL:U00039; NID:g466582; PIDN:AAB18457.1; PID:g466618
A;Experimental source: strain K-12, substrain MG1655
R;Zhao, S.; Sandt, C.H.; Feulner, G.; Vlazny, D.A.; Gray, J.A.; Hill, C.W.
J;Bacteriol. 175, 2799-2808, 1993
A;Title: Rhs elements of Escherichia coli K-12: complex composites of shared and unique
A;Reference number: A36902; MUID:93259920; PMID:8387990
A;Accession: B36902
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1397-1411 <ZHA>
A;Note: sequence extracted from NCBI backbone (NCBIN:132073, NCBIP:132075)
R;Sadosky, A.B.; Davidson, A.; Lin, R.J.; Hill, C.W.
J;Bacteriol. 171, 636-642, 1989
A;Title: rhs gene family of Escherichia coli K-12.
A;Reference number: A91901; MUID:89123133; PMID:2644231
A;Accession: A30092
A;Molecule type: DNA
A;Residues: 1221-1411 <RES>
R;Feulner, G.; Gray, J.A.; Kirschmann, J.A.; Lehner, A.F.; Sadosky, A.B.; Vlazny, D.A.;
J;Bacteriol. 172, 446-456, 1990
A;Title: Structure of the rhsA locus from Escherichia coli K-12 and comparison of rhsA w
A;Reference number: I54935; MUID:90094253; PMID:2403547
A;Accession: I54935
A;Status: translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-100 <SAD>
A;Cross-references: GB:M29717; NID:g147622; PID:g147623
C;Comment: the rhs core consist of two distinct parts: a large N-terminal core that is c
C;Genetics:
A;Gene: rhsB
A;Map position: 77 min
C;Superfamily: rhsF protein
C;Keywords: transmembrane protein
F;1-26/Domain: signal sequence #status predicted <SIG>
F;27-1411/Product: rhsD protein #status predicted <MAT>
F;28-55/Domain: transmembrane #status predicted <TMM>

Query Match   3.8%;   Score 340.5; Pred. No. 2, 2.6e-10;
Best Local Similarity 20.4%; Mismatches 459;   Indels 490; DB 2;   Length 1411;
Matches   287;   Conservative 172; Mismatches 459;   Indels 490;   Gaps   70;

Qy   418 KTETW-RSADNSYRSEITETTFDESGNPLTKVI------KDKKTQKIIS----PS------PS------ 461
          | | : || :||| : || | | | |
Db   218 RTQTFHREAAGEFSGEITGVT-DGAWRHFRLVLTTQAQRAEEARQQAISGGTEPSAFPDT 276

Qy   462 ------------THWEYYPPAGEVDNCPPEP---YGFT------RFVK 488
                       | || : || | ||:|
Db   277 LPGTEYGRDNGIRLSAVWLTHDPEYP----ENLPAAPLVRYGWTPRGELAVVYDRSGK 331

Qy   489 KIIQTPYDSEFK--------DDPEKFIQYRYSLIGSQHVTLKIEERHYSATQLLNSTL 539
          :| || :: | :: || :
Db   332 QVRSFTYDDKYRGRMVAHRHTGRPE--IRYRYD--SDGRVTEQLNPAGLSYT------ 379

Qy   540 FQYNTDKSELGRLLKQTEC--TKGENGKTYSVVHKFTYTKQDDTLQQSHSITTHDNFTIH 597
          | | : || | | | | :::| || | |
Db   380 YQYEKDRITITDSLDRREVLHTQGEAG-LKRVVKK------------EHADGSVT 421

Qy   598 RSQVRSRYTGRLFSDTDTKDIVTQMSYDKLGRLLTRTLNSGTPYANTLTYDELNNLQDD 657
          ::| || : || || || | ::   :
Db   422 QSQFDA--VGRLRAQTDAAGRTTEYSPDVVTGLITRIT---TPDGRASAFYYNHHNQ--- 473

Qy   658 NRPPVITTTDVNGNQLRNEFDGAGRHVSQCLKDSD-----GDGKFYTIHTQQYDEQGRH 712
          :: | | : | || ::| : || : || ||
Db   474 ----LTSATGPDGLELRREYDELGRLIQETAPDGDITRYRYDNPHSDLPCATEDATGSR 528

Qy   713 HTSTYSDYLTNGRQQTDPDKVHLLSMSKSYDNWGQIANTHWSYGVSEKITVDPI-TLIATK 771
          ||:||: || || | | || |||| || | ||| | :| :| || | | |
Db   529 KTMTWSRY--GQLLSFTDCSGYVTRYDHDRFGQMTAVHREEGLSQYRAYDSRGQLIAVK 585

Qy   772 QLQSNSNNVQTGKEVTYTPSQOPIQTLFDEAGHLQ---------SCHLTRDGWDRVRKE 824
          :|:|: ||| |:| || ||| | || | | :| |
Db   586 DTQGHETRYE------------------------------------YNIAGDLTAVIAPDGSRNGTQYDAWGKAVRT 626

Qy   825 TDAIGQCTIYQYDNYNRVIQITLPDGTIVNRKYAPFSDTLITDIRVNGISLGQQTF-D 882
          | || | ||:|: ||| |:|| | :|| | ||| || | | | ||
Db   627 TQG-GLTRSMEYDAAGRVIRLTSENGS---HTTFRYDVLDRLIQETFDGRTQRYHHD 680

Qy   883 GLSRLTQSQDGGRVWAYTYSAGNDQCFSTVITPDGQFIHYQYQPELDDAVLQVASNEITQ 942
          ||| | | | : | |: : : | | || | | ::   ::| : |
Db   681 LTGKLIRSEDEGLV------------THWHYD-EADRLTHRTVKGETAE 716

Qy   943 QFSYNPVTGAL--LKAVAEGQSLTPIY---PSGRL--KMENINDMKKMSYLW------ 988
          :::| | || | | | ||: || : | ||:: |
Db   717 RWQYDE-RGWLTDISHISEGHRVA-VHYRYDEKGRLTGERQTVHHPQTEALLWQHETRHA 774

Qy   989 -TLRGLENGYTDLTGTIQKISRDTHGRVTQIKDSSIKTLNVDDLNRHIGSQVTDLATGH 1047
           : | |||| |:| | | |:| :| ||| | ||| : | | | | | :
Db   775 YNAQGLAN--RCIPDSLPAVEWLTYG------SGYLAGMKLGD------------ 809

Qy   1048 MLTTTVEF--DGLNREIGRKLCDSSGHTLDIQQSWLKTQQLANRIVKLNGVLQRTEQYSY 1105
          ::||: :   : : :: | : :   :: ||:| |: || ||| ||::|
Db   810 --TPLVEYTRDRLHRETLR------SFGRYELTTAYTPAGQLQSQ--HLNSLLS-DRDYTW 859

Qy   1106 DSRNRLNQYKCDGAECFTDKYGHSIVTQNFTYDIYGNITACHTTFADGTEDHATFKFANP 1165
          ||:|| | | || | :| | | | :::   : | :
Db   860 NDNGELIR------ISSPRQ------TRSYSYSTGRLTGVHTTAA-----NLDIRIPYA 902

Qy   1166 ----NH 1202
          : ||
Db   903 TDPAGNRLPD--PELHPDSTLSMWPDNRIARDAHYLYRYDRHGRLTEKTDLIPEGVIRTDD 961

Qy   1203 GNTENFTYDTLGRLQNGQ----------GSVGYDPLNRLVSQKTDTLDCEL-- 1244
          | | : | | | | | |: |
Db   962 ERTHRYHYDSQHRLVHYRTQYEEPLVESRYLYDPLGRVAKVWRRERDLTGWMSLSRK 1021

Qy   1245 ---YY---RETMLVNE----VRNGEMIRLLR----TGETIIAQQRA---- 1276
             || : |: | : :|| |||: | |||
Db   1022 PQVTWYGWDGDRLITIQNDRSRIQTIYQPGSFTPLIRVETATGELAKTQRRSLADALQQS 1081

Qy   1277 --------------SKVLLTGTDSQQSVILTSDKQNLSQEA------------------ 1303
                        |:: | | | | | :|| ::|:
Db   1082 GGEDGGSVVFPPVLVQMLDRLESEILA---DRVSEESRRWLASCGLTVEQMKNQMDPVYT 1138

Qy   1304 --------------------YSAYGKHKSTAND--ASILGYNGERADP 1329
                               : || | :| |:||||| |
Db   1139 PARKIHLYHCDHRGLPLALISTEGATAWCAEYDEWGNLLNEENHQLQQLIRLPGQQYDE 1198

Qy   1330 VSGVTHLGNGYRSYDPTLMRFHTPDSLSPFG-AGGINPYSYCLGDPINRSDPSGHLSWQA 1388
          | | | | | | | | | | | | || || | |
Db   1199 ESGLYY--NRHRYYDPLQGRYITQD---PIGLKGGWNLYGYQL-NPISDIDPLGLSMWED 1252

Qy   1389 WTGIGMGIAGLLLTIATGMAIAAAGIAAAIASTSTTALAFGALSVTSDITSIVSGALE 1448
                                                  :::
Db   1253 -----------------AKSG---ACTNGLCGTLSA--------MIGPDKFDSIDSTAY 1283

Qy   1449 DASPKASSILGWSVSMGMGAAGLAESAIKGGTKLATHLGAFAEDGENALLKSTSESSRIKW 1508
          |
Db   1284 DALNKINS------------------------------QSICEDKEFA---- 1301

Qy   1509 GVTRSLDREIVRNEEGQVIKDHSRGYTDNFMGKGEQ------AILVHGDKDGFLYHTEG 1561
                                                     | | |: || |: ||
Db   1302 ----------------GLICKDNSGRYFSTAPNRGERKGSYPFNSPCPNGTEKVSAYHTHG 1346

Qy   1562 NKHNGK--GPYTRHTPEQLVDYLKDNNI 1587
          ||::|| ||| |||| || || |||
Db   1347 ADSHGEYWDEIFSGKDEKIVK-SKDNNI 1373

RESULT 13
H64780
rhsD protein precursor - Escherichia coli (strain K-12)

```
C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence revision 17-Sep-1997 #text_change 09-Jul-2004
C;Accession: H64780; JS0625; B30092; I69401; S16026
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: H64780
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1426 <BLAT>
A;Cross-references: UNIPROT:P16919; GB:AE000156; GB:U00096; NID:g1786705; PIDN:AAC73599.
R;Experimental source: strain K-12, substrain MG1655
R;Sadosky, A.B.; Gray, J.A.; Hill, C.W.
Nucleic Acids Res. 19, 7177-7183, 1991
A;Title: The RhsD-E subfamily of Escherichia coli K-12.
A;Reference number: JS0625; MUID:92115567; PMID:1766678
A;Accession: JS0625
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-378,'A',380-1166,'G',1168-1426 <SAD>
A;Cross-references: EMBL:X60999; NID:g42732; PID:g42733
R;Experimental source: strain K-12
R;Sadosky, A.B.; Davidson, A.; Lin, R.J.; Hill, C.W.
J. Bacteriol. 171, 636-642, 1989
A;Title: rhs gene family of Escherichia coli K-12.
A;Reference number: A91901; MUID:89123133; PMID:2644231
A;Accession: B30092
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1232-1426 <RES>
A;Cross-references: GB:M29719; NID:g147644; PIDN:AAA24541.1; PID:g147645
C;Comment: the rhs core consist of two distinct parts: a large N-terminal core that is c
C;Genetics:
C;Gene: rhsD
C;Superfamily: rhsF protein
C;Keywords: transmembrane protein
F;1-26/Domain: signal sequence #status predicted <SIG>
F;27-1426/Product: rhsD protein #status predicted <MAT>
F;28-55/Domain: transmembrane #status predicted <TMM>
A;Accession: I69401
A;Residues: 1-100 <SA2>
A;Cross-references: GB:M21764; GB:J04224; NID:g147646; PIDN:AAA24542.1; PID:g147649
R;Feulner, G.; Gray, J.A.; Kirschmann, J.A.; Lehner, A.F.; Sadosky, A.B.; Vlazny, D.A.;
J. Bacteriol. 172, 446-456, 1990
A;Title: Structure of the rhsA locus from Escherichia coli K-12 and comparison of rhsA w
A;Reference number: I54935; MUID:90094253; PMID:2403547
A;Accession: I69401
A;Molecule type: DNA
```

```
Query Match        3.8%;    Score 339;  DB 2;   Length 1426;
Best Local Similarity  20.2%;    Pred. No. 3.2e-10;
Matches  257;   Conservative 150;   Mismatches 409;   Indels 458;   Gaps  60;
```

```
Qy  572 KFTYTKQDDTLQQSHSITTHDNFTIHRSQVRSRYTGRLFSDTDTFKDIVTQMSYDKLGRLL 631
            :||||: .: ||:     :: ||:..  |||  .|| ||:|
Db  317 RYTYTEAGELL----AVYDRSNTQVRAFTYDAQHPGRMVAHRYAGRPEMRYRYDDTGRVV 372

Qy  632 TRTLNSGTPYANTLTYDYELNNLQDDNRPPFVITTD------IIVTDSLNRREVLHTEGGAGLKRVVKKE 668
            |: .: .:                :|     ||          :
Db  373 EQLNPAG------LSYRY----LYEQDR------ITVTDSLNRREVLHTEGGAGLKRVVKKE 418

Qy  669 -VNGNQLRNEFDGAGRHVSQCLKDSDGDKFYTIHTQQYD-----EQRHHTSTYSDYL 721
           :|| :||| | :|| | |      ::| .|||   | |:|     |:|  :
Db  419 LADGSVTRSGYDAAGRLTAQ---TDAAGRTEYGLNVVSGDITDITTPDGRETKFYYND-- 474

Qy  722 TNGRQQT---DPDKVHLSMSKSYDNWGQIANTHWSYGVSEKITVDPITLTATKQLQSNSN 778
            ||:          :||    : :  :   ::   |:  :.|  ::   :||: |:: :
Db  475 -GNQLTAVVSPD--GLESRREYDBPGRLV------------------------SE 502

Qy  779 NVQTGKEVT-TYTPSQQPIQITLFDEAGHLQSCHTLTRDGWDRVRKETDAIGQCCTIYQD 837
            : ::|:|| :|  |                 :::: |  | |::  :: :||||
Db  503 TSRSGETVRYRYDDAHSELPATTTDATG---STRQMTWSRYGQLLAFTDCSGYCTRYEYD 559
```

```
RESULT 14
E85509
hypothetical protein Z0268 [imported] - Escherichia coli (strain O157:H7, substrain EDL9
```

```
Qy  838 NYNRVIQITLPDGTIVNRKYAPFSTDTLITDIRVNGISLGQQTPDGLSRLTQSQDG-GRV 896
            ::| ||:   |   ||:|    |::||:|  |:||||:|| || :||| ||:|| |||
Db  560 RFGQM------------TAVHRE----------------EGISL-YRRYDNRGRLTSVKDAQGRE 595

Qy  897 WAYTYSAGNDQCPSTVITPDGQFIHYQYQ- 925
            |: :|| .:|| : |:  |||||||||
Db  596 TRYEYNAAGDL--TAVITPDGNRSETQYDAWGKAVSTTQGGLTRSMEYDAAGRVISLTNE 653

Qy  926 --------PELDDAVLQVASNEITQQFSYNPVTGALLKAVAEGQSLTPIYYPSGRLKME 976
                      :     :  :| .| | .|: :   |:  |:  |  | :  .: ||:|
Db  654 NGSHSVFSYDALDRLVQQGFDGRTQRYHD-LTGKLTQSEDEGLVILWYDESDRITHR 712

Qy  977 NINDMKKMSYLWTLRGLENGYTDLTGTIQKISRDTHGRVTQIKDSS----IKTTLNYDDL 1032
            :|:  : ::   : ::: ||    :|  |: ||   :|  :::
Db  713 TVNGEPAEQWQY------------------DGHGWLTDISHLSEGHRVAVHYGYDDK 751

Qy  1033 NRHIG--SQVTDLATGHML 1049
            ||:|.   | ||||.|
Db  752 GRLTGECQTVENPETGELLWQHETKGHAYNEQGLANRVTPDSLPPVEWLTYGSGYLAGMKL 811

Qy  1050 --TTTVEF--DGLNREIGRKLCDSSGH--TLDIQQSWLKTQQLANR----- 1089
            :: :::|    | ::  |  | ::: :   |::: ::: || |:
Db  812 GGTPLVEYTRDRLHRETVRSFGSMAGSNAAYELTSYTPAGQLQSQHLNSLVYDRDYGWS 871

Qy  1090 --IVKLNGVLQRTEQYSYDSRNRLNQYKCDGAEC-----PTDKYGHSI- 1130
            ::| :.|:|:   :: :: :::|::::| |:|      |::|| :
Db  872 DNGDLVRISGPRQ-TREYGYSATGRLESVRTLAPDLDIRIPYATDPAGNRLPDPELHPDS 930

Qy  1131 ------VTQN---FTYDIYGNITACHTTFADG---TEDHATFKFANPTDPCQLTEV 1174
                    :: | : ::::  :   ::   |::
Db  931 TLTVWPDNRIAEDAHYVRHDEYGRLTEKTDRIPAGVIRTDDERT---------- 975

Qy  1175 HHTHPDMPDNIRLKYDKAGRVINITD-NHGN--TENFTYDTLGR------------ 1215
            |||                :|:|:| ::   :   | ::: ||
Db  976 HHYH--------YDSQHRLVFYTRIQHGEPLVESRYLYDPLGRRMAKRVWRERRDLTG 1025

Qy  1216 ----LQNGQGSVYGYDPLNRLVSQKTDTLDCELYR--------- 1247
            :|: : ::|| ||| :::||   ::| |  |||:|
Db  1026 WMSLSRKPEVTWYGWDG-DRLTTVQTDTRIQTVVEPGSFTPLIRVETENGEREKAQRRS 1084

Qy  1248 --ETMLVNEVRNG-------EMIRLL-RTGETIIA------------QQRASKV 1279
             : :  :  :|:         ::|:: :|||: :            |||:.:|
Db  1085 LAETLQQEGSENGHGVVFPAELVRLLDRLEEIRADRVSSESRAWLAQCGLTVEQLARQV 1144

Qy  1280 LLTGTDSQQS---------------VILTSDKQNLSQEA-YSAYGKHKSTANDASILGYN----- 1323
            :|:|:||::               | : ::|:|:|.:|  :|:|| :|::  :  :|:
Db  1145 EPEYTPARKAHLYHCDHRGLPLALISEDGNTAWSAEYDEWGNQLNEENPHHV--YQPYRL 1202

Qy  1324 -GERADPVSGVTHLGNGYRSYDPTLMRFHTPDSLSPFG-AGGINPYSYCLGDPINRSDPS 1381
            |: :| | .|| :  ||| ||    |  |  |:   |    ||  | | |  |: |:|
Db  1203 PGQQHDEESGLYY--NRHRYYDPLQGRYITQD----PMGLKGGWNLYQYPL-NPLQQIDPM 1256

Qy  1382 GHLSWQAWTGIGMGIAGLLLTIATGGMAIAAAGGIAAAIASTSTTALAFGALSVTSDITS 1441
            ||| :     ::
Db  1257 GLL--QTWDDARSG-----ACTGGV---CGVLSRIIGPSKFDSTADAALD---------- 1296

Qy  1442 IVSGALEDASPKASSILGWVSMGMGAAGLAESAIKGGTKLATHLGAFAEDGENALLKSTS 1501
Db  1297 ----ALKETQNRS-------LCNDMEYSGIVCKDTNG---------KYFASKAETDNLR--K 1336

Qy  1502 ESSRIKWGVTRSLDREIVRNEEGQVIKDHSRGYTDNFMGKGEQAILVHGDK--DGFLYHT 1559
            |:|::| :| :  :  |:  :
Db  1337 ESYPLKRKCPTGTDRVAAYHTHG---ADSHGDYVDEFFSSSDKNLVRSKDNNLEAFYLAT 1393

Qy  1560 EGNKH--NGKGPY 1570
            |:||    |||| |
Db  1394 PDGRFEALNNKGEY 1407
```

```
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:Accession: E85509
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E85509
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1404 <STO>
A:Cross-references: UNIPROT:Q8XED9; GB:AE005174; NID:g12512977; PIDN:AAG54537.1; GSPDB:G
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z0268
C:Superfamily: rhsF protein

Query Match      3.8%;      Score 336.5;   DB 2;    Length 1404;
Best Local Similarity   20.0%;    Pred. No. 4.3e-10;
Matches  252;   Conservative  130;   Mismatches  373;   Indels 505;   Gaps  58;

Qy   376  YHL-LTSBCKQQNGYIQTETAYYAIIGHNFDSQPSQFQLPKT-KTETWRSADNSYRSEI  433
Db   245  FHLVLTTQAQRAEVFRKQRATSLSSPAGPR--SASSSLVFPDTLPAGTEYGADNGIRLEA  302

Qy   434  TETTFDES------GNPLTKVIKDKKTQKIISPSTHWEYYPPAGEV----DNCPPEPYGF  483
Db   303  VWLTHDPAYPDELPAAPLARYT------------YTASGELRAVYDRSGTQVRGF  345

Qy   484  TRFVKKIIQTPYDSE----------FKDDPEKFIQYRYSLIGSQSHVTLKIEERHYSATQL  534
Db   346  A----------YDAEHAGRMVAHHYAGRPES--RYRYDDTG------RVTEL  379

Qy   535  LN---STLFQYNTDKSELGRLLKQTEC--TKGENGKTYSVVHKFTYTKQDDTLQQSHSI  588
Db   380  VNPEGLDYRFBEYGQDRVTITDSLNRREVLYTEGEGG-LKRVVKK------  422

Qy   589  TTHDNFTIHRSQVRSRYTGRLFSDTDTKDIVTQMSYDKLGRLLTRTLNSGTPYANTLTYD  648
Db   423  -EHADGSITRSE------------------YDEAGRLKAQT------  444

Qy   649  YELNNLQDDNRPPFVITTDVNGNQLRNEFDGAGRHVSQCLKDSDG------DGKFYT  700
Db   445  ------------DAAGRRTEYSLHMASGAVTAVTGPDGR--T  472

Qy   701  IHTQQYDEQGRHHTSTYSDYLTNGRQQTDPDDKVHLSMSKSYDNWGQIANTHWSYGVSEKI  760
Db   473  VR-YGYNSQRQVTSVTYPDGLRSSREYDEKGRLAAETSRSGE------TTRYSYD-----  520

Qy   761  TVDPITLTATKQLQSNSNNVQTGKEVTYTPSQOPIQITLFDEAGHLQSCHTLTRDGWDR  820
Db   521  -DP------ASELPTGIQDATGSTKQM------------AWSR  544

Qy   821  --VRKETDAIGQCTTIYQDNVNRVIQITLPDGTIVNRKYAPFSTDTLIITDIRVNGISLG  877
Db   545  YGQLLTFTDCSGYTTRYEYDRYGQQIAVHREEGISTYSSYNP------RG  588

Qy   878  QQTFDGLSRLTQSQDGGRVWAYTYSAGNDQCPSTVITPDGQFIHYQYQPELDDA---VLQ  934
Db   589  Q------LVSQKDAQREYRYEYSAAGDL--TAIVAPDGSRSEIQY---DAWGKAVS  634

Qy   935  VASNEITQQFSYNPVTGALLKAVAEGQSLTPIYYPSGRLKMENINDMKGMSYLWTLRGLE  994
Db   635  TTQGGLTRSMGYDAAGRITVLTNENGSQSTFRYDPVDRLTEQRGFDGRTQRY------  686

Qy   995  NGYTDLTGTI-----------QKISRDTHGRVTQIKDSS-  1022
Db   687  --HYDLTGKLTQSEDEGLVTLWHYDASDRITHRTVNGDPAEQWQYDEHGWLTTLSHTSEG  744

Qy  1023  --IKTTLNYDDLNRHIGSQ--VTDLATGHML  1049
Db   745  HRVSVHYGYDDKGRLTGERQTVENPETGEMLWEHETGHAYSEQGLATRQEPDGLPPVEWL  804

Qy  1050  -------TTTVEF--DGLNREIGRKLCDSSGHTLDIQQSWLKTQQLANRIVKL  1093
Db   805  TYGSGYLAGMKLGGTPLVEYMRDRLHRETARSF--GGEAYELATAWNTSGQLRSRHLNL  861

Qy  1094  NGVLQRTEQYSYDSRNRLNQYKCDGAEBCTDKYGHSIVTQNFTYDIYGNITACHTTFADG  1153
Db   862  P---QLDRDYDWNDNGQL-IRISGPQ-----ESREYRYSDTGRLTGVHTTAANL  906

Qy  1154  TED-HATFKPANP-TDPCQLTEVHHTHPD------MPDNIRL-----KYDKAGRVIN  1197
Db   907  DIDIPYATDPAGNRLPDP------ELHPDSTLTAWPDN-RIAEDAHYVYRYDEYGRLAE  958

Qy  1198  ITD------NHGNTENPTYDTLGRL----QNGQGSV---YGYDPL-----  1229
Db   959  KTDRIPEGVIRMHDERTHYHYDSQHRLVFHTRIQHGEPQVESRYLYDPLGRRTGKRVWR  1018

Qy  1230  --------NRLVSQKTDTLDCELYR----ETMLVNEVRNGEMI  1261
Db  1019  RERDLTGWMSLSRKPEETWYGWDGDRLTTVQTQGTRIQTVYQPGSFTPLLRIETENGEQA  1078

Qy  1262  RLLR------TGETIIAQ----QRASKVLLTGTDSQQS  1289
Db  1079  KARHRSLAEVLQEDTGVTLPAELAVMLGRLERELRQGSVSEESQQWLAQCGLTAEQMAAQ  1138

Qy  1290  ------------VILTSDKQNLSQEAYSAYGK--HKSTANDASILGYN  1323
Db  1139  LEAEYIPERKLHLYHCDHRGLPLALISPGGETAWQGEYDEWGNLLGETSAQHLQQSLRLP  1198

Qy  1324  GERADPVSGVTHLGNGYRSYDPTLMRFHTPDSLSPFG-AGGINPYSYCLGDPINRSDPSG  1382
Db  1199  GQQYDEESGLYYNRNRY--YDPLQGRYITQD----PIGLBGGWNLYQYPL-NPIEHIDPLG  1252


RESULT 15
C65159
rhaA protein precursor - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 09-Jul-2004
C:Accession: C65159; S47814; I69402
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: C65159
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1377 <BLAT>
A:Cross-references: UNIPROT:P16916; GB:AE000437; GB:U00096; NID:g2367249; PIDN:AAC76617.
A:Experimental source: strain K-12, substrain MG1655
R:Plunkett, G.
submitted to the EMBL Data Library, March 1994
A:Accession: S47814
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-864,'M',866-1377 <PLU>
A:Cross-references: EMBL:U00039; NID:g465582; NID:g465731
A:Experimental source: strain K-12, substrain MG1655
R:Feulner, G.; Gray, J.A.; Kirschmann, J.A.; Lehner, A.F.; Sadosky, A.B.; Vlazny, D.A.;
J. Bacteriol. 172, 446-456, 1990
A:Title: Structure of the rhsA locus from Escherichia coli K-12 and comparison of rhsA w
A:Reference number: I54935; MUID:90094253; PMID:2403547
A:Accession: I69402
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-864,'M',866-1377 <RES>
A:Cross-references: GB:M29716; NID:g147611; PIDN:g147614
A:Experimental source: strain K-12
C:Comment: the rhs core consist of two distinct parts: a large N-terminal core that is c
C:Genetics:
A:Gene: rhsA
```

```
C;Superfamily: rhsF protein
C;Keywords: transmembrane protein
F;1-26/Domain: signal sequence #status predicted <SIG>
F;27-1377/Product: rhsD protein #status predicted <MAT>
F;28-55/Domain: transmembrane #status predicted <TMM>

Query Match      3.7%;    Score 331.5;  DB 2;    Length 1377;
Best Local Similarity  20.8%;  Pred. No. 7.8e-10;
Matches  288;  Conservative 170;  Mismatches 450;  Indels 477;  Gaps  75;

Qy     211  LFPGQKEGY-RTELRFLNRQLNSIHNFSLGNENPLTWSFGYTPIGKNGILGQWITSMTAP  269
                  ||| |   || :| |:  |:  |||     | ||: |  | ||
Db     126  LFPGE-DGYSRSESLWLVR--GGVAKLDEGHRLAALWQ-------------------     160

Qy     270  GGLKETVNYS--------NNNQG------HHFPQSANL--PVLPYTLMKQVPGAGQ-     310
                  |: ||             :  ||         ||:
Db     161  -ALPEELRLSPHRYLATNSPQGPWWLLGWCERVPEADEVLPAFLPPYRVLTGLVDRFGKT  219

Qy     311  PAIQAEYSYTSHNYVGGSNGIWNNKLDNLYGLMTEYNYGSTESRRYKDKEGHDQIVRIE  370
                  :| |:     :    |                :|:::|
Db     220  QTFHREAAGFSGEITGVTDGAWRH------FRLVL-----TTQAQR-----         255

Qy     371  RTYNYNHLLTSECKQQNGYIQTTETAYYAIIGHNFDSQPSQFQLPKT---KTETWRSADN  427
                                 |: |:| |           :  |
Db     256  -------AEEARQQ---------AISG---GTEPSAF--PDTLPGYTEYGR--DN     287

Qy     428  SYRSEITETTFDESGNPLTKVIKDKKTQKIISPSTHWEYYPPAGEVDNCPPEP---YGFT  484
                  :| |                          |  |||          :|:|
Db     288  GIRLSAVWLTHDPE-----------------YP----ENLPAAPLVRYGWT     317

Qy     485  ----------RFVKKIIQTPYDSEFK---------DDPEKFIQYRYSLIGSQSHVTLKIE  525
                      |||: | ||          | ||   | |:||
Db     318  PRGELAVVYDRSGKQVRSFTYDDKYRGRMVAHRHTGRPE--IRYRYD---SDGRVTEQLN  372

Qy     526  ERHYSATQLLNSTLFQYNTDKSELGRLLKQTEC--TKGENGKTYSVVHKFTYTKQDDTLQ  583
                  :  :|          | :|| |     |:  |  |  :| |: |  |  | ::|
Db     373  PAGLSYT-------YQYEKDRITITDSLDRREVLHTQGEAG-LKRVVKK-------     413

Qy     584  QSHSITTHDNFTIHRSQVRSRYTGRLFSDTDTKDIVTQMSYDKLGRLLTRTLNSGTPYAN  643
                  |  :  | :     |:  | |: : : |  :      |
Db     414  -------EHADGSVTQSQFDA--VGRLRAQTDAAGRTTEYSPDVVTGLITRIT----TPDGR  462

Qy     644  TLTYDYELNNLQDDNRPPFVITTDVNGNQLRNEFDGAGRHVSQCLKDSD------GDGKF  698
                  : :||            |   |:|      | | | ||:     |  :      |||
Db     463  ASAFYYNHHNQ------------LTSATGDGLELRREYDELGRLIQETAPDGDITRYRYDNPH  514

Qy     699  YTIHTQQYDEQGRHHTSTYSDYLTNGRQQTDPDKVHLSMSKSYDNWGQIANTHWSYGVSE  758
                  |   |  : || :| ||  :|  | :|   ||| |:         |  |
Db     515  SDLPCATEDATGSRKTMTWSRY--GQLLSFTDCSGYVTRYDHDRFGQMTAVHREEGLSQ  571

Qy     759  KITVDPI-TLTATKQLQSNSNNVQTGKEVTYTPSQQPIQITLFPDEAGHLQ-------SC  810
                  |                  | |  |   | |    :| |      ||
Db     572  YRAYDSRGQLIAVKDTQGHETRYE--------------YNIAGDLTAVIAPDGSR     612

Qy     811  HTLTRDGWDRVRKETDAIGQCTIYQDYNYNRVIQITLPDGTIVNRKYAPFSTDTLITDIR  870
                  |    |    : :      |  :| |   |  : :         | |
Db     613  NGTQYDAWGKAVRTQG-GLTRSMEYDAAGRVIRLTSENGS-----HTTFRYDVLDRLIQ  666

Qy     871  VNGISLGQGTF--DGLSRLTQSQDGGRVWAYTYSAGNDQCPSTVITPDGQFIHYQYQPEL  928
                  |           |    | ||| | |:    |       :  |       | |
Db     667  ETGFDGRTQRYHHDLTGKLIRSEDEGLV------------THWHYD-EA     702

Qy     929  DDAVLQVASNEITQQFSYNPVTGAL--LKAVAEGQSLTPIYY--PSGRL-KMENINDM  981
                  :| |  :|:: :  : ||| :|   ||: | : |   |:     ||  |  :|:| |
Db     703  DRLTHRTVKGETAERWQYDE-RGWLTDISHISEGHRVA-VHYRYDEKGRLTGERQTVHHP  760

Qy     982  KKMSYLW-------TLRGLENGYTDLTGTIQKISRDTHGRVTQIKDSSIKTTLNYDDLN  1033
                  |: |||       |      ||  | |:       | |:       | | |::|
Db     761  QTEALLWQHETRHAYNAQGLAN--RCIPDSLPAVEWLTYG------SGYLAGMKLGD--  809

Qy    1034  RHIGSQVTDLATGHMLITTVEF--DGLNREIGRKLCDSSGHTLDIQQSWLKTQQLANRIV  1091
                  |:   |        : |      | |   |  :      |   :| : |||| || :
Db     810  -----------------------TPLVEYTRDRLHRFTLR----SFGRYELTTAYTPAGQLQSQ--  846
```

```
Qy    1092  KLNGVLQRTEQYSYDSRNRLNQYKCDGAECPTDKYGHSIVTQNFTYDIYGNITACHTTFA  1151
                  ||  |:       | :     :             |:|  | |   :||:|| || |
Db     847  HLNSLLS-DRDYTWNDNGELIR------ISSPRQ------TRSYSYTTGRLTGVHTTAA   893

Qy    1152  DGTEDHATFKFANPTDPC--QLTEVHHTHPD----MPDN------IRLKYDKAGRVIN  1197
                  |   ||  |||   |||      | : |  |    |: |
Db     894  ----NLDIRIPYATDPAGNRLPD-PELHPDSTLSMWPDNRIARDAHYLYRYDRHGRLTE  947

Qy    1198  ITD--------NHGNTENFTYDTLGRLQNGQ--------GSVVGYDPLNRLVSQKTDT  1239
                  |:|        :    | |:|  | |               ||  | |  |   | |
Db     948  KTDLIPEGVIRTDDERTHRYHYDSQHRLVHYTRTQEBPLVESRYLYDPLGRRVAKRVWR  1007

Qy    1240  LDCEL-------YY----RETMLVNE-----------VRNGEMIRLLR---TGETI  1270
                  |  |:       ||     : ||  |            | ||
Db    1008  RERDLTGWMSLSRKPQVTWYGWDGDRLTTIQNDRTRIQTIYQPGSFTPLIRVETATGELA  1067

Qy    1271  IAQQRA--------------------SKVLLTGDSQQSVIL------------TS  1294
                  ||| |:                    |:  : | |  ||:
Db    1068  KTQRRSLADALQQSGGEDGGSVVFPPVLVQMLDRLESEILADRVSEESRRWLASCGLTVE  1127

Qy    1295  DKQN------------------LSQEA-----------YSAYGKHKSTAND---AS  1318
                  || |                  || |           ||  |: :| |
Db    1128  QMQNQMDPVVTPARKIHLYHCDHRGLPLALLSKEGTTEWCAEYDEWGNLLNEENPHQLQQ  1187

Qy    1319  ILGYNGERADPVSGVTHLGNGYRSYDPTLMRFHTPDSLSPFG-AGGINPYSYCLGDPINR  1377
                  | ||  |     ||      |     :: |   |       |    | ||: | | |
Db    1188  LIRLPGQQYDEBSGLYY--NRHRYRYDPLQGRYITQD--PIGLKGGWNFYQYPL-NPVTN  1241

Qy    1378  SDPSG  1382
                  :|| |
Db    1242  TDPLG  1246

Search completed: February  9,  2005,  18:18:45
Job time : 78 secs
```

```
         GenCore version 5.1.6
         Copyright (c) 1993 - 2005  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          February 9, 2005, 18:14:07 ; Search time 182 Seconds
                                              (without alignments)
                                              2993.957 Million cell updates/sec

Title:           US-09-889-874A-23
Perfect score:   8879
Sequence:        1 VYIKFLKLFRRITMSDNNEF.........PRKIILGRTEKTVKPKTFRP 1673

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5
Searched:        1373511 seqs, 325702437 residues
Total number of hits satisfying chosen parameters:     1373511

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published Applications AA:*
  1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
  2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
  3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
  4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
  5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
  6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
  7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
  8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
  9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep:*
 17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
 18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep:*
 19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
 20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	932.5	10.5	982	15	US-10-365-742-106	Sequence 106, App
2	517.5	5.8	1317	15	US-10-369-493-11243	Sequence 11243, A
3	420.5	4.7	2364	14	US-10-156-761-7834	Sequence 7834, Ap
4	395	4.4	820	14	US-10-156-761-7990	Sequence 7990, Ap
5	392	4.4	1250	14	US-10-156-761-7572	Sequence 7572, Ap
6	391.5	4.4	2386	14	US-10-156-761-7751	Sequence 7751, Ap
7	386	4.3	2234	15	US-10-282-122A-46565	Sequence 46565, A
8	361	4.1	1510	15	US-10-418-861B-55	Sequence 55, Appl
9	353.5	4.0	843	15	US-10-282-122A-76214	Sequence 76214, A
10	350	3.9	1515	15	US-10-282-122A-47600	Sequence 47600, A
11	342	3.9	1397	15	US-10-282-122A-43059	Sequence 43059, A
12	340.5	3.8	1411	15	US-10-282-122A-43060	Sequence 43060, A
13	339	3.8	1426	9	US-09-912-020-340	Sequence 340, App
14	339	3.8	1426	14	US-10-282-122A-42617	Sequence 42617, A
15	331.5	3.7	1377	15	US-09-815-242-10384	Sequence 10384, A
16	331.5	3.7	1377	9	US-10-287-274-467	Sequence 467, App
17	331.5	3.7	1377	14	US-10-282-122A-42731	Sequence 42731, A
18	328	3.7	1572	15	US-10-282-122A-69415	Sequence 69415, A
19	322	3.6	1385	15	US-10-282-122A-68242	Sequence 68242, A
20	317.5	3.6	1565	15	US-10-282-122A-44558	Sequence 44558, A
21	314	3.5	1395	15	US-10-282-122A-50586	Sequence 50586, A
22	307.5	3.5	932	15	US-10-282-122A-44617	Sequence 44617, A
23	306.5	3.5	1530	15	US-10-282-122A-68175	Sequence 68175, A
24	298.5	3.4	1531	15	US-10-282-122A-49308	Sequence 49308, A
25	297	3.3	1512	15	US-10-282-122A-78386	Sequence 78386, A
26	291.5	3.3	1253	15	US-10-282-122A-68914	Sequence 68914, A
27	291	3.3	2346	15	US-10-072-012-491	Sequence 491, App
28	290.5	3.3	2802	9	US-09-808-602-81	Sequence 81, Appl
29	290.5	3.3	2802	10	US-09-800-198-69	Sequence 69, Appl
30	290.5	3.3	2802	15	US-10-072-012-489	Sequence 489, App
31	289.5	3.3	1364	15	US-10-282-122A-75309	Sequence 75309, A
32	285	3.2	2613	15	US-10-038-854-42	Sequence 42, Appl
33	285	3.2	2628	15	US-10-038-854-40	Sequence 40, Appl
34	285	3.2	2721	15	US-10-038-854-38	Sequence 38, Appl
35	285	3.2	2725	15	US-10-038-854-36	Sequence 36, Appl
36	284.5	3.2	2715	15	US-10-042-865-52	Sequence 52, Appl
37	284.5	3.2	2715	15	US-10-029-020-51	Sequence 51, Appl
38	280	3.2	1317	9	US-09-815-242-5118	Sequence 5118, Ap
39	280	3.2	1317	15	US-10-282-122A-43495	Sequence 43495, A
40	279.5	3.1	944	15	US-10-282-122A-47806	Sequence 47806, A
41	278	3.1	985	15	US-10-282-122A-43061	Sequence 43061, A
42	276.5	3.1	1438	15	US-10-282-122A-78360	Sequence 78360, A
43	273.5	3.1	1688	15	US-10-144-194A-113	Sequence 113, App
44	273.5	3.1	1737	9	US-09-808-602-83	Sequence 83, Appl
45	273.5	3.1	1737	10	US-09-800-198-71	Sequence 71, Appl

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

ALIGNMENTS

```
RESULT 1
US-10-365-742-106
; Sequence 106, Application US/10365742
; Publication No. US20030204868A1
; GENERAL INFORMATION:
;    APPLICANT: Collmer, Alan
;    APPLICANT: Alfano, James R.
;    APPLICANT: Cartinhour, Samuel W.
;    APPLICANT: Schneider, David J.
;    APPLICANT: Tang, Xiaoyan
;    TITLE OF INVENTION: PSEUDOMONAS AVR AND HOP PROTEINS, THEIR ENCODING
;    TITLE OF INVENTION: NUCLEIC ACIDS, AND USE THEREOF
;    FILE REFERENCE: 19603/4112
;    CURRENT APPLICATION NUMBER: US/10/365,742
;    CURRENT FILING DATE: 2002-02-12
;    PRIOR APPLICATION NUMBER: 60/356,408
;    PRIOR FILING DATE: 2002-02-12
;    PRIOR APPLICATION NUMBER: 60/380,185
;    PRIOR FILING DATE: 2002-05-10
;    NUMBER OF SEQ ID NOS: 209
;    SOFTWARE: PatentIn Ver. 2.1
;    SEQ ID NO 106
;    LENGTH: 982
;    TYPE: PRT
;    ORGANISM: Pseudomonas syringae pv. tomato DC3000
US-10-365-742-106

  Query Match             10.5%;  Score 932.5;  DB 15;  Length 982;
  Best Local Similarity   29.7%;  Pred. No. 5.7e-56;
  Matches  285;  Conservative  142;  Mismatches  357;  Indels  175;  Gaps  33;

Qy        642 ANTLITYDYELNNLQDDNRPPFVITTTDVNGNQLRNEFDGAGRHVSQCLKDSDGDGKFYTI 701
                :::||| ||::|  :|   |||||| :| ||||  | :|| | ||          :
Db         31 SSTLGYRYD-----DWNQR--CCTTTDDN-VQTYEYSDPIGSDVHK-------GPIQKT 74

Qy        702 HTQQYDEQGRIHTSTYSDYLTNGRQQTDPDKVHLSMSKSYDWWGQIANTHWSYGVSEKIT 761
```

```
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 11243
;   LENGTH: 1317
;   TYPE: PRT
;   ORGANISM: Methanosarcina mazei
;   FEATURE:
;   NAME/KEY: unsure
;   LOCATION: (1)..(1317)
;   OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-11243

Query Match            5.8%;  Score 517.5;  DB 15;  Length 1317;
Best Local Similarity 21.2%;  Pred. No. 1.1e-26;
Matches  334;  Conservative 194;  Mismatches 505;  Indels 541;  Gaps 72;

Qy   39 TGLYNIQITLGHIVGNGNLGPTLPLTL--SYSPLNKTDIGFGIG--FNFGLSVYDRKNSL 94
          :| ::: |    :  :| |  |:||         |:|  ||  |   |||| ||
Db   56 TGSYFYQYQDLSIPGRG-----LPLTVSRSYNSMDNRSGLFGSGWTFNYNMNLTVDNNGN 110

Qy   95 LSLSTGENYK--VIETDKTVKLQQKKLDNLRFEKDLKENCYRIIHKSGDIEVLTGFNNN 151
          ::  |                     :| :
Db  111 VTVLGGDGHDTDTYILNPDGTYSRPLSVFDDL----------IKNSDETYTLTYKKDQT 157

Qy  152 AFPDLKVPKKLLN--PAGHAIYIDWNFEATQPRLNRIYDDLDGHDIPLLNLEYQGLIKTI 208
          ||| |                   :: ::     |: ||
Db  158 KYNFSSEGKLVNIVDKNGNQI----NFTYTGEQLTKV-TDASGREL-ILAYDHNG----- 206

Qy  209 LTLFPGQKEGYRTELRFLNRQLNSIHNFSLGNENPLTWSFGYTPIG------KNGILGQ- 261
                        :: ::     :                             :|
Db  207 ----------------------HIISITDPMGRVWSVSYDDQQNLIQCKNPIGGKLS 241

Qy  262 ------WITSMTARGG--LKETVNYSNNNQGHHFPQSANLPVLPYVTLMKQVPGAGQP 311
               :  :         |:|||:   :         :
Db  242 YTYDENHWMTSITDPRGNIPMKNTY-------DEKGHVISQS-------- 276

Qy  312 AIQAEYSYTSHNYVGGGSNGIWNNKLDNLYGLMTEYNYGSTESRRYKDKEGHDQIVRIER 371
                  : || :|:           ::||||
Db  277 ----------------NSLNATY-----TFNYDS-ENRK----------------- 293

Qy  372 TYNNYHLLTSECKQONGYIQTTETAYYAIIGHNEFDSQPSQFQLPKTKTETWRSADNSYRS 431
                               :|||:
Db  294 ----------------------TTETD-------------PFGKNKTYSFDEHFW--- 313

Qy  432 EITET------------TFDESGNPLTKVIKDKTQKIISPSTHWEYYPPAGEVDNCPPEPYG 482
          |:|:|          :
Db  314 ELNETNQLGYTISAYDENGNRISVTNENSKTTKLA------YDANGNIIK-TTNPLG 364

Qy  483 FTRFVKKIIQTPYDSEFKDDPEKFIQYRYSLIGSQSHVTLKIEERHYSATQLLNS----- 537
          ::: |           :|| ||
Db  365 YSK-------SMTYDS-----------KNNLISQTDELRHKTSFEYDDNSNLIKSIDALG 406

Qy  538 --TLFQYNTDKSELGRLLKQTECTK------------GENGK---TYSVVHKFT 574
            |:| |:||      |:|            |||  : ||||:
Db  407 HETVFSYD----KYGQVIGETDSNKKTATFSYNNNGDQITITDANGKTSAFTYDTVGRVT 462

Qy  575 YTKQDDTLQQSHSITTHDNFTIHRSQVRSRYTGRL-------FSDTDTKDI 618
          |:|||||    :  : |:| :
Db  463 TKTDAKGNRYTFQYDALDNLLSITD-------PYA-------PMGQTTSNTMLLEIKLVLQMLKVDXQSI 516

Qy  619 VTQMSYDKLGRLLTRTLNSGT--------PYA-------NTLTYDYEINNLQDDNRPPFVIT 665
          ::: ||||||:|   : :              :: |: :
Db  517 LTTLXISXLKXRMQWVEXLGTNMMPLAIWFPRQTQKGHKTSYDYDPLNRQ-------VS 568

Qy  666 TTDVNGNQLRNEFDGAGRHVS-------------QCLKDSDGDGKFYTIHTQQYDE 708
          :| :|| :   :| ||| :               ::     | |: :| : |
Db  569 VTNALGKTTRNKYDAIGNKISITNAYGKSTRYSYNSLNQLVKVTNAMGK---VVRYNYDA 625

Qy  709 QGRHHTSTYSD--------YLTNGRQQTDPDKVHLSMSKSYDNWGQIANTHWSYGVSEKIT 761
          ::| |: :  |        :  ||  ||   : :
```

```
Db   75 WKQSGDPEGR---------------------------------WLNL-------FGKPDRIR 103
          || :| |:|                                  :| |        ::|| :
Qy  762 VDPITLRATKQLQSNSNNVQTGKEVTVTPSQQPIQITLFDEAGHLQSCHTLLTRDGWDRV 821
          | :|| | |                                  ::: :||
Db  104 ----TLTAGKTGRSRTHSMSRSRNLTT--TEQEL------SRQTFLYDGLGRC 144

Qy  822 RKETDAIGQCTIYQYDNYNRVIQITLPDGTIVNRKYAPFSTDTLITDIRV---NGIS--- 875
          :: ||  |  |::| |:|:|  |: ||||::  :| |||:|: :  ::        |:
Db  145 TEQRDALQQSTLFSYDNWSRMVSSTLADGSVINRSYAPQSSSELATMLEVVHQNGTTRTV 204

Qy  876 LGQQTFDGLSRLTQSQDGGRVWAYTYSAGNDQCPSTVITPDGQFIHYQYQPELDDAVLQV 935
          || | ||||:|:||::||||:  :|| :|  |   |:|||| |  |||| ||: | ||::
Db  205 AGTQKFPDGLERVTQTKTGDRVEQFNYDAGEMQ-PRSRTTAGLDNINFPYTRALTDQIFSS 263

Qy  936 ASNEITQQFSYNPVTGALLKAVAEGQSLTPIYYPSGRLKMENINDMKKMSYLW-----T 989
          :: :|:  | || ::|  ||: ::|:  ||  | ||   | : : |  :|   :     :
Db  264 TAPDETAKFDYDKTSARLIEATNPQGTRTYRYRYDVHNQLTGETWDNL--LGQAWETRHQSS 321

Qy  990 LRGLENGYTDL-------TGTIQKISRDTHGRVTQIKDSSIKTTLNYDDLNRHIGSQVTDLA 1044
          | |   | :|        || |:|  |:| |||  |: :|:||::||| | |: | ||: |:
Db  322 LLGRPIKRTDLKKGEAAGAETRYDYDTLGRIRFINQSNLRTTIDYDVLGQLCKVATEDLQ 381

Qy 1045 TGHMLTTTVEFDGLNREIGRKLCDSSGHTLDIQQSWLKTQQLANRIVKLNGVLQRTEQYS 1104
          :| :: :  |:|  ::|| | | :: |:||:  |:| | ||   ||  |   ||:  :||
Db  382 AGTGVIIDMEYDDQGQEILRTQTASNQAALTLTQTWAVDGLLKTRDLQQAGSPLLHETFS 441

Qy 1105 YDSRNRLNQYKCDGAECPTDKYGHSIVTQNFTYDIYGNITACHTTFADGTEDHATFKFAN 1164
          ||  :||  :  | || | |: |  ::  :|  ||:||: |  | |:|||:: |  | | |
Db  442 YDPRGRLTLVNYLGSSLPRDELQREMTRQIFSPDELDNITLCQTRFTDGTSERAAFKYGS 501

Qy 1165 P------TDPCQLTEVHHTHPDMPDNIRLKYDKAGRVINITDNGNTENFTYDTLGRL--- 1216
                 ||| ||  :  : | :|:  :: | |||  :: |::|:   |: |||:  ||
Db  502 PGDDKHKDRCQLLSIAYTPPRKTPDPTFSYDANGN--QLKDEHGN--SLHYDSQSRLLQV 557

Qy 1217 -QNGQG--SVYGYDPLNRLVSQKTDTLDCEL--YYRETMLVNEVRNGEMIRLLRTGETII 1271
           : | |  |:| |  |||:|::  |  :  | : ::|  :|:  :||| :|  | |: |
Db  558 AETGGAPISQYRYDGHNQLVATR-DGNESEILRFYEGHQLSSTVQEDRTQYILHLGEQPL 616

Qy 1272 AQQ--RASKVLLTGTDSQQSVILTSDKQNLSQEAYSAYGKHKSTANDASILGYNGERAD 1328
          |:|   | :|::|:||::||: ::|| :  |||  | |||:: |: :|::| :||| :
Db  617 GQQIVDDAEQTLLLLTDANQSVMGEFQQGVRKAVYSAYGERHSEEALLSTAGFNGEVRE 676

Qy 1329 PVSGVTHLGNGYRSYDPTLMRFHTPDSLSPFGAGGINPYSYCLGDPINRSDPSGH----- 1383
          | :|   |||||::|:| ||||| ||:|||| | |:||| |||| || | || |
Db  677 AANGWYLLGNGYRAYNPLLMRFHSPDFLSPFAEGGVNPYTYCLGNPIALRDPTGHDASGQ 736

Qy 1384 ------LSWQ-----AWTGIGMGIA----GLLLTIATGGMAIAAAGGI----------- 1416
               | |      |::|||:||:    ||:|: ||| |:| | |:|
Db  737 TGRLRRPDEGALPMQQGGGDIMGWVGVGIGVVFTVLGVAATIATLGTATPVTGPVTVLGI 796

Qy 1417 ------AAAIASTSTTALAFG----ALSVTSDITSIVSGALEDASPKASSILGWVS---- 1462
                :||:|||: :|:: ||    |||: : :|: |:||| || :|: ||::|| |
Db  797 SMTASAAAAVSTVSTGALIVGTALTAASTTANTVAIVNN-----DQTAGEVGGWLGIAAV 851

Qy 1463 ----MGMGAAGLAESAIKGGTKLA----------THLGAFAEDGENALLKSTSESSR 1505
              :| ||:|:|  :::  ||:           ::||:|| | ||| | ||:|::|
Db  852 PVGLVGFGAGAVVARAVARAVAAAAKVAAANAGTIGVRSVSRIG-LAAAGARRTISSAASSAR 909
```

```
RESULT 2
US-10-369-493-11243
; Sequence 11243, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
;   APPLICANT: Cao, Yongwei
;   APPLICANT: Hinkle, Gregory J.
;   APPLICANT: Slater, Steven C.
;   APPLICANT: Goldman, Barry S.
;   APPLICANT: Chen, Xianfeng
;   TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;   TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
;   FILE REFERENCE: 38-10(52052)B
```

```
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7834
; LENGTH: 2364
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-7834

Query Match         4.7%;   Score 420.5;   DB 14;   Length 2364;
Best Local Similarity   21.6%;   Pred. No. 1.8e-19;
Matches 320;  Conservative 190;  Mismatches 550;  Indels 423;  Gaps 73;
```

```
Qy    134  RIIHKSGDIEVLTGFNNNAFDLKVPKKLLNPAGHAIYIDWNPEATQPRLNRIYDDLDGHD  193
           :  ||
Db    702  RKTYGDGDDIVLPVINFNRTLD-----NKVGSA--ELNF--PRIKEIHGDL------    744

Qy    194  IPLLNLEYQGLIKTILTLPGQKEGVRTELRFLNRQLNSIHNFSLGNENPLTWSFGYTPI   253
           :  ||                                             : ||
Db    745  --------GATTTVSYGFAN-----ACIDHLPAQAS-------NTQDCYWQ-KWTPE    781

Qy    254  GKNGILGQWI-------------------TSMTAPGGLKETVNYS-NNNQGHHFPQSANLPVLP-   297
           :  :
Db    782  GETDSKTGWFKKFLVTQVQVDPTVTTNQDGAPVMTSYTYEDGAGWHP---TNDPLIKDE   838

Qy    298  ------YVTLMKQV-PGAGQPAIQAEYSYTSHNYVGGSNGIWNNKLDNLYGLMTEY      347
               ||
Db    839  DESWTDWRGYQEVQVTTGAGAGQK-------TKKSWLYRGLSGDRTSKAD-------     881

Qy    348  NYGSTESRRYKDKEGHDQIVRIERTYNNY--------------HLLTSECKQQNGYIQTTETAYYA   399
           ||                                         : :||
Db    882  -ASATKTVTVDDDG---------------NNYTDDSWLSGRILSTSLRDDTG--TSHERTYHK    927

Qy    400  IIGHN----FDSQPSQFQLPKTKTETWRSADNSYRSEITETFPDES---------GNPLTKVI   449
               ||                                                 ||
Db    928  YWDHNTAQYDGLPDARFVRBKETTTNTKVSSGWREHTVETEYDDTEGASTTFGLPMR----     984

Qy    450  KDKKTQKIISPSTHWEYYPPAGEVDN-CPPEPYGFTRFVKKIIQTPYDSEFKDDPEKFIQ   508
           |  |                                          : ||
Db    985  TDDWGQSSVS-----------------DNRC---------------TTYGRAYNTD------   1008

Qy    509  YRYSLIGSQSHVTLKIEERHYS--------------ATQLLNSTLFQYN---------      543
           ||                                  :  :
Db   1009  -NYDSTGAQRWTVLQDQVKHYSVGCSSIADSKQDGYASTLYDNATSVDANKPVDGNPTES   1067

Qy    544  ------------------TDKSELGRLLKQTE----------------CTKGENGKTYSVVHKFTVTKQDDTLQQ   584
                             ||
Db   1068  RTYTKSGSYRSTWSGYDDAGRVMWSDGSHNRTLTKYSPANTWPLNGVIVTTPDPDGALP   1127

Qy    585  SHSITTHDNFTIHRSQVRSRYTGRLFSDTDTKDIVTQMSYDKLGRLL-----------    631
           : :                                            ||
Db   1128  AHTALTSTAWT--------SRFWGKPTSIQDANGNVTKMSLDAAGRLLVEVWRPTETGSSPS   1180

Qy    632  --------------------------------TRTLNSGTPYANTLTYDYELNNLQDDNRPPFVI    664
Db   1181  MKFSYTIPTSTNSAGVPDAVDGYPHVATHVLQSGTTYLSSHAYVDGLGRAR-ETQTPMGN   1239

Qy    665  TTTDVNGNQLRNEFDGAGRHVSQCLKDSDGDGKFVTIHTQQYDEQGR----------HH    713
Db   1240  GVDAATGNEVPN------RQVSVTRYDSAGN---VTGTSAVFRNQGTAGSGPSSAKVED    1290

Qy    714  TSTYSDYLTN--GRQQTDPDKVHLSMSKSYDNWGQIANTHWSYGVSEKITVDPITLTATK    771
           ||                                               : :||
Db   1291  LPSYTDLVLDWAGRAITS------RLQVNGASQDAGRVDTTYDGDFTSVKNKVDAAADTYTD   1346

Qy    772  QLQSNSNNVQTGKEVT-------TYTPSQQPIQITLFDEAGHLQSCHTLTRDGWDRVRKETD    826
           ||                                            : :||
Db   1347  VYGQVSKVVEHTGSSATYTTAYTTAKDELIKIT--DPRGN-DTSYTYDWAGRQRTATDDSD   1403

Qy    827  AIGQCTIYQYDNYNRVIQITLPDG-TIVNRKYAPFSTDTLITDIRVNGISLGQQTFDGLS    885
           ||                                                    : :||
Db   1404  AGVSST--EYDKNGQISKITSNGGKTVLDIYGYDALGRKTA---VRSGDTELAAWTWDGLN   1458
```

```
Db    626  VGNLISTTDENGHKINYGYDSLNRQVSVTDALRKTTRNKYDAVGNKISITNAYGKSTRYS    685

Qy    762  VDPITLTATKQLQSNSNNVQTGKEVTTYTPSQQPIQITLFDEAGHLQSCHTLTRDGWDRV    821

Db    686  Y-----NSLN-----QLVKVTDAMGGVVRNYDAVGNLIS----------------------    715

Qy    822  RKETDAIGQCTIYQYDNYNRVIQITLPDGTIVNRKYAPFSTDTLITDIRVNGISLGQQTF    881

Db    716  --TTDANGRKTNYGYDSLNRQVSITNALGKTTRNKYDAVGNKISSTDANWR--LTKYSY     770

Qy    882  DGLSRLTQSQDG-GRVWAYTYSA-GNDQCPSTVITPD--GQFIHQY-----QPELDDAV    932

Db    771  DSLNRLVKVTDAMGGVVRRYTYDAVGN-----LISTTDAKGHKTDYGYDSLDRQVSITDPL   825

Qy    933  LQVASNBITQQFSYNPVTGALLKAVABGQSLTPIYYPSGRLKMENINDMKKMSYLWTLRG    992

Db    826  GRTARNK------YDAVGNKISSTDEDGKTTSYGYDVLNRLTKVSYPDDQKVSY-------   873

Qy    993  LENGYTDLTGTIQKISRDTHGRVTQIKDSSIKTTLNYDDLNRHI------GSQV--TDLA   1044

Db    874  ------------NYDAVGNRLTMKDSHGTTAYKYDKLNRLLSVLNPDGQKVSYTYNK     918

Qy   1045  TGHML---------TTTVEFDGLNREIGRKLCDSSGHTLD--IQQSWLKTQQLANRIVKL   1093

Db    919  VGNRVKGMTYPDGKTTSYSYDAVNRLIG--VIDSDGHITSYSYAXNGNLKTMNTPNGV---   973

Qy   1094  NGVLQRTEQYSYDSRNRLNQYKCDGAECPTDKYGHSIVTQNFTYDIYGNITACHTTFADG   1153

Db    974  ------KTE-YSYDKANRL----VELINKNTTQVVSSYKYTLDAAGNRLKVDEQLAEG    1020

Qy   1154  TED---------HATFKFANPTDPCQLTEVHHTHPDMPDN-IRLKYDKAGRVINIT-    1199

Db   1021  VESGDSELKESQLLTTTYGY---DKLYRLTKV------DYPSNKTVSYKYDSMGNRISMTT   1072

Qy   1200  --DNHGHTENFTYDTLGRLQNGQGSVYGYDPLNRLVSQKTDTLDCELY-YRETMLVNEVR   1256

Db   1073  NVDGIGSTISYKYDAADQLLQSGNISYSYDKKGNLIKKRVNSTQFMSYSYDE--ANRLK    1129

Qy   1257  N----------GEMIRLLRT--GETIIAQQRASKVLLTGDSDQQSVILT    1293

Db   1130  NVSEFVSNTNTPKSSYNFEYDGDGNRIIKTTINGE----NAQSTKYVLDINSALPQVLTE   1185

Qy   1294  SDKQNL--------SQEA-----YSAYGKHKSTANDASIL-------------------   1320

Db   1186  SDTKNTTCYTYGTDLISMTNSENAEYYYHYDGLGSVRSLSDSKGIIKNTYLYDAFGQVQK   1245

Qy   1321  ------------GYNGERADPVSGVTHIGNGYRSYDPTLMRFHTPDSLS--PFGAGGINPYSY   1369

Db   1246  EIGTVDNEFRFTGEQMDDETGLIYLRARY--YDPSVGRFITKDVIKGRRVTTQSINRVYY   1303

Qy   1370  CLGDPINRSDPSGH   1383

Db   1304  TTNNPVNLVDLTGY   1317
```

```
Db          26  DTEGKITVFRYDAWGRVDKIT-----------------------|------TAEGRVTIVFTYD-DANRVTISMLRGTGFNS :   74
                     .:||:    .:.|:                                       .: ||| .     :

Qy         560  KGENGKTYSVVHKFYTKQDDTLQQSHSITTHDNFTIHRSQVRSRYTGRLFSDTDTKDIV  619
                                                                                  :||
Db          75  DGHTGPTW------------------TYAYTSDSVTAAGTTATDPET-HATKYQHDSDGQVSDVTDAMGHK  127

Qy         620  TQMSYDKLGRLLTRT--LNSGTPYANTLTYDYEL-NNLQDDNRPPVITTTDVNGNQLRN  676
                     :|||:.|||:||||        : : ||||       ||:|:
Db         128  RSTKFDANHSIDTSTDAMGSGTTPGNVTDYGFNTRNNLE-----------TITQPTGKTVN  178

Qy         677  EFDG-AGRHVSQCLKDSDGDGKFYTIHTQQYDEQGRHHTSTYSDYLTNGRQQTDPDKVHL  735
                     ::|: .      : :                   :                   : |
Db         179  HWQTIAGADVPKDSTNPDGEKTDFT-----YDAVGN-----------------------TM  211

Qy         736  SMSKSYDNWCQIANTHWSYGVSEKITVDPITLTA--TKQLQSNSNNVQTGKEVTTYTPS  792
                     ::.|:  .: .|     :: |                 ||
Db         212  SVAQTGTGGGNVSYTY------------------NPASPTCCGFAGQRCTEKTKMTAAKTVT-----  256

Qy         793  QQPIQITLPDEAGHLQSCH----------TLTRDGWDRVRKETDAIGQCTTIYQYDNVN--RV  842
                     : |:||:    ::                |. |
Db         257  ----TFHYDSAGNLDTVTPPAPLAKITYTYDALGRTKTVTDARGVTVTYTYDNRDRTRI  311

Qy         843  IQITLP--------DGTIVNRKYAPFSTDTLITDIRVNGISLGQOTFDGLSRLT--QSQ  891
                     :|.||            :: ||:       :                |:
Db         312  VDTTNKARVEYWYDGDGNLTQR-----TDSTGTIKYD------------FDPLQRETIRTLQ  356

Qy         892  DGGR-VWAYTYSAGNDQCPSTVITPDGQFIHYQYQPELDDAVLQVASNEITQQFSYNPVT  950
                     |||.                                  :||.
Db         357  DGSQTLLAY-------------------------DPAG-----------TVDYTWNEV-  388

Qy         951  GALLKAVAEGQSLTPIYYPSGRLK--MENINDMKKMSYLWTLRGLENGYTDLTGTIQKIS 1008
                                ::|:                        : :||               : | |:
Db         389  --------NKLAELKDPAGRVTTYKYNNNDVRTTT----TYPG-----------GTVQKVD  426

Qy        1009  RDTHGRVTQIKDSSIKTT-----------LNYDDLNRHIGSQV---TDLATGHMLTTTVEFDGL 1058
                     :|||||          :::                              : |
Db         427  PDNSSRPKTIKTTSPKGTPVDLAYSYGYGTSGTTEGSKIRSSTDNVTG--LKTTYTYDGA  484

Qy        1059  NREIGRKLCDSSGHTLDIQQSWLKTQQLANRIVKLNGVLQRTEQYSYDSRNRLNQYKCDG 1118
                                                                     :.                    :
Db         485  GRFSYAE--EKKGTTLN--SSWLYCYDLAGNL--------------TSQCTAAGCPGATTYTVND  531

Qy        1119  AECPTDKYGHSIVTQNFTYDIYGNITACHTTFADGTEDHATFKFANPTDPCQLTEVHHTH 1178
                     :||:|:|:     |:::|                         :::|       |
Db         532  AQQLTAKNGS---TTNWSYDKIGNETAGAST-PEGTRTAEKW-----TDHSQLTSL--  578

Qy        1179  PDMPDNIRLKYDKAGRVINITDNHGNTENFTYDTLG--RLQNGQQSVYGYDPLNRLIVSQK 1236
                     : .|:        |    :. :   :                        |
Db         579  -------TVGGK--TYTGQYGSTDQSERIRLGDTYFHNGPIGLAG-----------  614

Qy        1237  TDTLDCELYYRETMLVNEVRNGEMIRLLRTGETIIAQQRASKVLLTGTDSQQSVILTSDK 1296
                     :: ||               : :||     .   |::
Db         615  TSTAGVDTGFNR--------------------EPGGTLNSMTTGGKAYCYLTDALGSVIALTDE  658

Qy        1297  --QNLSQRAYSAYGKHKSTANDASILGYN--GERADPVSGVTHLGNGYRSYDPTLMRFHT 1352
                     :| |:|:    ||.    |:|:|                                |
Db         659  TGAKVNTYAYSPRGVTRASTAESKNPQPYRFAGGYQD--VTGLYHYGARY--YDPNIGRFNS  715

Qy        1353  PDSLSPFGAGGINPYSYCLGDPINRSDPSGHLSWQAWTGIGMGIAGLLLTIATGGMAIAA 1412
                     ||:| |           |:||.      ::  |            :|:
Db         716  PD---PSCQEE-NPYLYABGDPVNRIDPNGLFSFSDALDTGSDIFGVVTGCLAGVTAAAE  771

Qy        1413  AGGIAAAIASTSTTALAFGALSVTSDITSIVSGALEDASPKASSILGWVSMGMGAAG 1469
                     ||||                       : ::. .:  ::|  .: :
Db         772  TGSIAYAAA-------------------------VGGVGAGVGSAVGSAVGLAVVGSCALG-GAAG  810
```

```
  Query Match             4.4%;  Score 395; DB 14; Length 820;
  Best Local Similarity  23.8%;  Pred. No. 2e-18;
  Matches  242; Conservative 120; Mismatches 375; Indels 280; Gaps  50;
```

```
Qy         501  DDPEKFIQYRYSLIGSQSHVTLKIEERHYSATQLLNSTLFQYNTDKSELGRLLKQTEC-T  559
```

```
Qy         886  RLTQSQDGGRVWAYTYSAGNDQCPSTVITPDGQFIHYQYQP-----ELDDAVLQVASNEI  940
                                                                              :|.|
Db        1459  -IT----GGK-GQITSTTSRDTNGNTYTTKTGEF-DVRGRPLKTTVTIPDTVKGLIAGS-Y 1510

Qy         941  TQQFSYNPVTGALLKAVAEGQSLTPIYYP-SGRLKMENINDMKKMSYLWTLRGLENGY--  997
                                                                                   :|:
Db        1511  TTGFTYD---------AADHVTSVAYPRAGGLPAETVN-----------ATYDGYGR 1547

Qy         998  --TDLTGTIQKISRDTHGRVTQIK------------DSSIKTTLNYDDIN--RHIGSQV 1040
                      :|||                          :|||            :     :|.||
Db        1548  PLRLQSALQTYIRSTGYDAYDRLTDRSYGVDSVLPGIGAAAQRTYSYDDSNGTRELKSVA 1607

Qy        1041  TDLATGHMLTTTVEPDGLNREIGRKLCDSSGHTLDIQQSWLKTQQLANRIVKL----NGV 1096
                      |||                                       ||  |:|| :| : :
Db        1608  T------TTT-----LNKVVSERQKDT--------------YTYDLAGKITELREQASGQ 1642

Qy        1097  LQRTEQYSYDSRNRL-NQYK----------CDGAECPTDKYGHSIVTQNFTYDIYGNITACHT 1148
                      |||              | |                |:||:|:  . |
Db        1643  TAQSCCFLYDDQARLTNAYTHTTTGICADKTKTASDFKGTAPYQTAYTYDRLGNLQSITN 1702

Qy        1149  TFADGT----------EDHATFKFANPTDPCQLTEV-HKTHPDMPDNIRLKYDKAGRVI 1196
                                           .||||                 |::|.||| |||
Db        1703  TDSAGTATLHDYLYPGYDDTGTWTTANANQPHGVRKIDHKTGSTTTKSDQLYYFDDGTMK 1762

Qy        1197  NITDNHGNTENFTYDTLGRLQ------NGQQSV---YGYDP----LNRLVSQKT----DTL 1240
                      |:||                        :      : |:               ::|
Db        1763  QRVE-PGITTDYTWTPQGQLEAVKTTKSSGSELTRYAYDADGNILVRTTPQETVASIDGM 1821

Qy        1241  DCELYYRETMLVNEV-RNGEMIRLLRTGETIIAQQRASKVLLTGTDSQQSVILTSDKQN- 1298
                               ||:|                        :||||
Db        1822  ELRTTNGTTVTATRVYASGTATVAMRTTEGTTATN--GKVTYLMADTQASTQLAVDASTG 1879

Qy        1299  -LSQEAYSAYGKHKS-TANDASILGYNGERADPVSGVTHLGNGYRSYDPTLMRFHTPDSL 1356
                        :: ||| .||||         | |||                           ||:
Db        1880  ASTRRRYTPFGDERSGTLPTGTDNGFLGKTEDTSTGLSLL--GARAYDPNLGRFLSPDPL 1937

Qy        1357  -SPFGAGGINPYSYCLGDPINRSDPSGHLSWQAWTGIGMGIAG 1398
                         ||: |     :||||:|| ||||||  |:
Db        1938  ATPYAPQNLNAYSAYATNNPISYSDPSG-LCRRDICGDGYPVAG 1979
```

```
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7572
; LENGTH: 1250
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-7572

Query Match        4.4%;     Score 392;    DB 14;    Length 1250;
Best Local Similarity  21.2%;  Pred. No. 6.4e-18;
Matches  271;  Conservative 187;  Mismatches 452;  Indels 368;  Gaps  66;

Qy   254 GKNGILGQWITSMTAPGGLKETVNYSNNNQGHHFPQ------SANLPVLFYVTLMKQVPG 307
             :  |: | ||| :  :  ::  : : ||:|
Db    59 GMEGVAEQGLRK-TARNAVEDAAEDASRQRGAKICEKDPVDVATGRMVLPQI-----DVSL 113

Qy   308 AGQPAIQAEYSYTSHNYVGGGSNGIWNNKLDNLY----------GLMTEYNYGSTES 354
             : |:  || :  |  |     ::  ||
Db   114 PGQLPLVVKRQFESSYRLGGWFGPTWSSTLDQRLEVDTAGVVLVGEDGLVLAYPH----PA 170

Qy   355 RRYKDKEGHDQIVRIERTYNNYHLLTSECKQQNGYIQTETAYYAIIGHNFDSQPSQFQL 414
                                                                 ||:|
Db   171 PGVPTLPSHGPRWSLDRTDGGYTL----TDQKSGHIR-------------------- 203

Qy   415 PKTKTETWRSADNSYRSEITETTFPESGNPLTKVIKDKKTQKIISPSTHWEYYPPAGEVD 474
                                                            :  ||:|
Db   204 --------HFVDRSDTLAVLEQLDDRNGNWIT----------------FEY------DAD 233

Qy   475 NCP---PEPYGFTRFVK----KIIQTPYDSEFKDDPEK-PIQYRYSLIGSQSHVTLKIEE 526
             |:         |  :  ||:|  |:|| :|  : :|
Db   234 GAPRYLTHSGGYRLRISTEAGRVTALHLASAAVDGGDQDLIRYGY----TDGHL------ 283

Qy   527 RHYSATQLLNST--LFQYNTDKSELGRLLKQTE----------CTKGENGKTYSV 569
             ::|:                                     ||:|
Db   284 ----TEVNSSGRPLQYVCD--ELGRIITSWTDTNDSHFSYAYDDEDRCTH-QSGAAGHL 335

Qy   570 VHKFYTYKQD------DTLQQSHSITTHDNFTIHRSQVRSRYTGRLFSDTDTKDIVTQM 622
             ||::|             : ::|| |
Db   336 RSTPAYGAVDPGTGAHTTVVTDSYGQTTH--YLINR------RCQVIAETDALGAVTRY 386

Qy   623 SYDKLGRLJLTRTLNSGTPYANTLTYDY-ELNNLQDDNRPPFVITTTDVNGNQLRNEFDGA 681
             ||:|                         ::  ||:|
Db   387 QRDRYNRLLSQT---DPLGHTTSFRYDDAGNLVAATRP-------DGREARAEYNAL 433

Qy   682 GRHVSQCLKDSDGDGKFYTIHTQQYDEQGRHHTSYSDYLTNGRQQTDPDKVHLSMSKSY 741
             :  |:                                             ::  ||:|
Db   434 GLPVK--LVNPDG----TITRQTFDERG---------NLTSVTDPS--GQTRFGY 472

Qy   742 DNWGQIANTHWSYGVSEKITVD----PITLT---ATKQLQSNSNNVQTGKEVTTYTPSQ 793
             ||:|                                              ::  ||:|
Db   473 DEGGRLTSMTDLGHTTGIVCDRAGLPLTVTDPLGAVTRYERDAFPGRTTA--ITDPTGAT 530

Qy   794 QPIQITLFDEAGHLQ------SCHTLTRDGWDRVRKETDAIGQCTIYQYDNYNRVIQIT 846
             ||:|                                                   ::  ||:|
Db   531 TRLEWTV---EGHLSRRTAPDGTSESWTYDGEGNCTSHTDPVGGVSLFEYTHFDLPTART 587

Qy   847 LPDGTIVNRKYAPFSTDTLITDIRVNGISLGQQTFDGLSRLTQSQDGGRVWAYTYSAGND 906
             ||:|                                              :  ||:|
Db   588 GPDGV----RY-EFEHD---TELR------LSQVTNPH--GLTWNYAYDAAGR 624
```

```
Qy   907 QCPSTVITPDGQFIHQYQPELDDAVLQVAS--NEITQQFSY-NPVTGALLKAVAEGQSL 963
             :  ||:|                                              : :|
Db   625 LVAETDF--DNRTLTYEY------DFAGRLASRTNALGQMIAFERNELGQIVRKDAAGQAT 677

Qy   964 TPIYYPSGRLKMENINDMKKMSYLWTLRGLENGYTDLTGTIQKISRDTHGRVTQIKDSSI 1023
             :  ||:|                                              ||:|
Db   678 TYAYDFTDQLAQA-------------------------TGPDGTALJILRDRHGRMISEAVNGR 716

Qy  1024 KTTLNYDDLNRHI----GSQVT----DLA----TGHMLTTTVEFDGLNREIGRKLCD 1068
             :  ||:|
Db   717 ALTYDYDELGRTRRTTPSSATTIWSDVAGRTSMTGSGRSIDFVYDEAGRELGRRF--774

Qy  1069 SSGHTLDIQQSWLKTQQLANRIVKLNGVLQRTEQYSYDSRNRLNQYKCDGAECPTDKYGH 1128
             :  ||:|
Db   775 --GH-ITLEHSFDALGRLTSQSV-LGPAGRRTQHRQY--------TYRAD--------GH 814

Qy  1129 SIVTQN-------FTYDIYGNITACHT------TF-ADGTEDHATFKFANP------- 1165
             :: :  ||:|
Db   815 LIGIEDQLSGKRRFDLDLAGRVTAVHAVNWTETYAYDAVGNQTSASWPAGHPGQEAIGNR 874

Qy  1166 ---TDPCQLTEVHHTHPDM-----------------PDNIRLKYDKAGRVINITDNHGNTEN 1207
             : :  ||:|
Db   875 TYTGTSLTRAGQVRYEHDGLGRIILRQKTRLSRKPDTWRYTWDAEDRLTSVTTPDGTRWR 934

Qy  1208 FTYDTLGR------------LQNGQGSVYGYDPLNRLVSQKTDTLDCELYYRETMLVNEVR---N 1257
             :||:|                                              : :|
Db   935 YTYDPLGRRTAKLRLAEDGETLV----------ERVDFTWDGTVLCEQTTTSPDLPHQVTLTWD 988

Qy  1258 GEMIRLLRTGETIIAQQRASKVL-------LTGTDSQQSVILTSDKQNLSQEAYSAY 1307
             :  ||:|
Db   989 HQGLRPVTQTERIVAADAPQQBIDSRPFAIVTDLVGTPSE---LLDERGEIWRTRATLW 1045

Qy  1308 GKHKSTANDASI--LGYNGERADPVSGVTHLGNGYRSYDPTLMRFHTPDSLSPFGAG-GI 1364
             :  ||:|
Db  1046 GSTTWAKNSTAYTPLRFPGQYYDPETGLHY--NYFRHYDPETARYVTPD--PLGLGPAP 1100

Qy  1365 NPYSYCLGDPINRSDPSG 1382
             ||  |   ||:|
Db  1101 NPAAY-VPNPHWAADPLG 1117
```

```
RESULT 6
US-10-156-761-7751
; Sequence 7751, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7751
; LENGTH: 2386
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-7751

Query Match        4.4%;     Score 391.5;    DB 14;    Length 2386;
Best Local Similarity  20.1%;  Pred. No. 1.9e-17;
Matches  379;  Conservative 214;  Mismatches 688;  Indels 601;  Gaps  93;
```

Qy   163 NP---AGHAIYIDWNFEATQPRLNRIYDDLDGHDIPLLNLEYQGLIKTILTLFPGQKEGY 219
          |||| :::|:: | | | : ::|:: |:|: |||
Db   698 NPDGTIGKTLWLDYIQRKT--------YGDGDDIVLPVINFNRTDL------------DNKVG- 740

Qy   220 RTELRFLNRQLNSIHNFSLGNENPLLTWSFG-----------------YTPIGKNG 257
          || : : |::
Db   1107 SRNRLN-QYK---------CDGAECPTDKYGHSIVTQNFTYDIYGNITACHTTFADGT----- 1154

Qy   258 ILGQWITS-MTAPGGLKEVNYSNNNQGHHFPQSANLPVLPYVTLMKQVPG--------AGQ 310
          ||:| : :: | ||| :|:| || :| |: || || |||
Db   1667 DQSRLTLAYHTTAGNCADTTKTTSDFKGTSPYQTGYTYDRLGNLQSVTDTNSAGAATTR 1726

Qy   311 PAIQAE---------MTEYNYGSTESRRYKD 359
          ||: | | : |||
Db   1727 DYLYPGYDDAGTWTTAN-----------ADQPHGVR-KINKVSAGTTPAAGTYTYYA 1771

Qy   360 KEGHDQIVRIERTYNNY--------HLLTSECKQQNGYIQTTETAYYAIIGHN----FDSQ 408
          ::: |: ||||
Db   1200 D---------NHGNTENFTYDTLGRL-------QNGQGSVYGYDPLNRLVSQKTDTLDCEL 1244

Qy   409 PSQFQLPKTETWRSADNSYRSEITETTFDES------------GNPLTKVIKDKKTQKIISPS 461
          |:| ||:|
Db   1772 DGAMKQRVEGSTTDYTWSRLGRLAATVKTTKTSGSDLTRTYDASGNLLVRTTP----- 1825

Qy   462 THWEYYPPAGEVDN-CPPEPYGFTRFVKKIIQTPYDSEFKDDPEKFIQRYSLIGSQSHV 520
          | :: : : :   :||| | :
Db   1245 YYRETMLIV----NEVRNGEMIRLLRT-------GETIIAQQRA-----SKVLLTGTDSQQS 1289

Qy   521 TLKIEERHYSATQLLNSTLFQYNTD--KSELGRILKQTECTKGENGKTYSVVHKFTYTKQ 578
          ||: |: | : | | : |: |:   :
Db   1826 --QETVASIGGTELRTTDGISATATRYYSFGATTVAMRTTDGNNTVNGKITYLMGDTQAS 1883

Qy   579 DDTIQQSHSITTHDNFTIHRSQVRSRYTGRLFSDTDTKDIVTQMSYDKLGRLLTRTLNSG 638
          :|:|| | : | || | ::| | : | :|| |:|
Db   1290 VILTSDKQ--NLSQRAYSAYGKHKS-TANDASILGYNGERADPVSGVTHLGNGYRSYDPT 1346

Qy   639 TPYANTLTYD---------YELNNLQDDN------------RPPFVI----TTTDVNGNQ 673
          | ||
Db   1884 TQIAVDAATGTATRRYYTPFGDERSGSLPTGTNHGFLGKTEDTNTGLSLL--GARAYDPS 1941

Qy   674 LRNEFDGAGRHVSQCLKDSDGDG---KF-YTIHTQQ-------YDEQGRHHTSTYS 718
          | | : | |
Db   1347 LMRFHTPDSLS--PFGAGGINPYSYCLGDPINRSDPSGSHLSWQA-------WTGIGMGIA 1397

Qy   719 DYLTN--------GR------QQTDPDKVHLSMSKS-----YDNWGQIA------ 748
          ||:   ::   ::
Db   1398 LGRFLSPDDPLSTPYDPQNLSAYSYSGNDPINYSDPSGLIKLNSDGTQCSDGWQKCGPGTV 2001

Qy   749 NTHWSYGVS-----------EKITVDPITLTATKQLQSNSNNVQTGKEVTTY------TP 791
          || ||
Db   1398 G-----------------------------LL------------LTIATGGMAIAAAG-GIAA 1418

Qy   792 SQQPIQITLFDEAG--------HLQSCHTLTRDGW---DRVRKETDAIGQCCTIYQYDNYNR 841
          || :::|||
Db   2002 GGGGDTPTPAPTVQDQLVDLLPREENGWDADRLAQVWVHYGQTTQGGGYWDAPVGDGDRT 2061

Qy   842 VIQITLPDGTIVNRKY---APFSTDTLITDIRVNGISLGQQTFDGLSRLTQSQDGG---R 895
          |: |||
Db   1419 AIAGTSTTAL--AFGALSVTTSDITSIV-----------SGALEDASPKASSILG 1459

Qy   896 VWAY----------------TYSAGNDQCPSTVITPDGQFIHYQYQPELDDAVLQVASNEITQQ 943
          || |:||
Db   2062 GMACFGRTACSEAFRVWKETHDFAKAKRVAATFCVENPKRCGADNGAYDSMKEASAVPI 2121

Qy   944 FSYNPVTGALLKAVA--EGQSLTPIYY--SGRLKMENI----NDMKKMSYLWTLRGLENG 996
          ||:   :
Db   1460 WVSMGMGAAGLAESAIKGGTKLATHLGAFAEDGENALLK--STSESSRIKWG--------V 1510

Qy   997 YTDLTGTIQKISRDTHGRVTQIKDSSIKTLN----------YDDLNRHIGSQ-VTDLATG 1046
          || ||
Db   2122 LLAGEMGAA------FSKVLKARGCSFKPTTRVLMKDGKTKPLGKIKPGDLVEAADP 2172

Qy   1047 HMLTTTVEFDGLNREIGRKLJCDSSGHTLDIQQSWLKTQQLANRIVKLNGVLQRTEQYSYD 1106
          ||:   :|
Db   1511 TRSLDREI-----------VRNEEGQVIKDHS----RGYTDNFMGKGEQAILV 1548

Qy   1288 GTAGSGGPSSPQPSDLPSYSDLVLDWAGRTVSSQIQVKGTPQKAGRVDTSYLGDYTSVTP 1347
          || |:
Db   2173 TSGHHREVREVTAVHLNHDDDLVDLSIRGLDGRIQTLHTTARHRIWDDTAQVWEQAGRLI 2232

Qy   1348 VDSTATDTYTDVYGQVSKVEHTASSAYTAYGYTAKGELAQITDPRGNNTLYTYDWAAAQ 1407
          || ||
Db   1549 HGDKDGFLYHTEGNCH---------NGKGPYTRHTPEQLVDYLKDNNIVDLTQGGDKP- 1597

Qy   1408 RKTTDDPDAGLSSSEYNENGQVSQTTATTNDVQTVLTYG--YDNLSRATSVRSGADELA 1464
          || ::
Db   2233 TGHK-----VNTSGNQHATITSVLAQRGAADMYDLITVEGLHTY---------YVLAGETPVL 2280

Qy   1465 AWWDDPAATGGKGQITSAVSRDASGNTYTTKIGKF------DERGRPLNTTVTLP 1514
          || ::|
Db   1598 VHLLSCYGKSS---------GAADKMAKYINRPVIAYSNKPTISQGLARIERKDFFLKSTYHS 1651

Qy   1515 TTVNGLAGDYTTSVTYDAADHITSVSYPAAGKLAAEKVTTYYDYGQPFRLTSSLG---- 1570
          || |
Db   2281 VHNGSCWSSTNRKTSVKNAFGHWKKH---------KSEFPNLNNAKEYVEAGTDFLRST--- 2330

Qy   1571 ---GTAYIDNTTYDAYGRLVE-RDYGAEFGGNGIQAQRQYGYDDSN--GTRWLRSIAT- 1622
          || :::::
Db   1623 ---TTTINDLVSEAQKDTYLYDNTGKLTELRE--------QASGQTAQSQCLRYD 1666

```
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
;
; SEQ ID NO 46565
; LENGTH: 2234
; TYPE: PRT
; ORGANISM: Bacillus anthracis
US-10-282-122A-46565

Query Match      4.3%;    Score 386;  DB 15;   Length 2234;
Best Local Similarity 18.7%;  Pred. No. 4.2e-17;
Matches 339;   Conservative 236;   Mismatches 633;   Indels 608;   Gaps 76;

Qy   21  FTQANNFTSAVSGGVDPR-------TGLYNIQITLGHIVG-----NGNL----------  57
Db  748  YSKVGHAEBSEISHEFVPRFSQDSGFIGMFDYWASIPVLNGKVNATNGNFIMSEKDITLSG 807

Qy   58  -GPTLPLTLSYSPLNKTDICFGIGFNPGLSV---YDRKNSLLSLS---------TGEN- 102
Db  808  RGPDVSVERTYNSQSKKVGLFGTGWSSGLEERVWADGNGNLLLISTDGANITFTRTGDNK 867

Qy  103  -----------YKVIETDKTVKLQQKKLDNLRFEKDLKENCYRIIHKSGDIEVL 145
Db  868  YQAPTGIYLEIKQVSGGYEIKDKDQTV--------------------TFYKSGDAQGR 905

Qy  146  TGFNNNAFDLKVP--------KKLLNPAGHAIYIDWNFBATOQPRLNRIYDDLDGHDIPL 196
Db  906  IEYTKDKYGNTTTYEYDGASRLSKVKNASGKELVL--QYDGNNKKAARVI----GPDNKT 959

Qy  197  LNLEYQGLIKTILTLFPGQ--KEGYRTELRFLNRQLNSIHNFSLGNENPLTWSFGY---- 250
Db  960  ITFNYDGDLLVSSTTPEGKVYKYGYD------NGVLTSIYDPQHTDAKPYKTSYAYENDR 1013

Qy  251  -----TPIGKNGIL----GQWITSMTAPGGLKETVNYSNNNQGHHFPQSANLPVLPYVTL 301
Db 1014  LVKVTDPLGKATTLAYNTGSKEVLITNPKGRKTVYTYND----------AGNPV------ 1057

Qy  302  MKQVPGAGQPAIQAEYSYTSHNYVGG-----GSNGIWNNKLDNLYGLMTEYN- 348
Db 1058  -KTVEDVGRLNLTTSYEYANANLVKTTTPKNQTETATYDGNGNVTSVTDEMGTEKFEYNK 1116

Qy  349  -YGSTESRRYKDKE--------GHDOIVRIERTYNNYHLLTSECKQQNGYIQTTETAYYAI 400
Db 1117  DNGIIKATDNEDRKTTVAYVGANTEVSQTDQGAN----TSSVIHHDQYGNPIETSKELS 1171

Qy  401  IGHNFDSQPSQFQLPKTKTETWRSADNSYRSEITETTFDESGNPLTKVIKDKKTQKIISP 460
Db 1172  AGGNLIQNPS-FEM--NGTEKWVKVDTNNSGSIS-----KDATPAPGGLGGESSLKITTK 1223
```

```
Qy  461  STH--WEYYPPAGEVDNCPPBPYGFTRFVKKII----------------QTPYDSEFKDDP 503
Db 1224  ATNNDWGYIAAIQEVTLEPNTYTLSGMVKTDLVNGAAFFNVQSLNENGAGIDGGWHDTR 1283

Qy  504  EKFIQYRYSLIGSQSHVTLKIEBRHYSATQLLNSTLFQYNTDKSELGRL-LKQTECYKGE 562
Db 1284  HNKVQGTSDWVNRQ--VTFKTTEQ----TRKVKIYLQVENGSATSSGSAWFDKIQLEKGE 1337

Qy  563  NGKTYSVHKFYTKQ--DDTLQQ-SHSITTH--------DNFTIHRSQVRSYTGRLFS 611
Db 1338  VSSSFNPVLNSSFEENWPDGFVPQWVRSSCSQHCERNDVSDDSFTGHSSIVMER-----S 1391

Qy  612  DTDTKDIVTQMSYDKLQGRLLTRTLNSGTPYANTLTYDYELNNLQDDNRPPFVITTTDVNG 671
Db 1392  EYGPNDI-----------GYRNRRVILNQKKAETVTLTAMSKSENVVND- 1428

Qy  672  NQLRNEFDGAGRHVSQCLKDSDGDGKFYTIHTQQYDEQGRHHTSYSDYLINGRQQTDP- 730
Db 1429  ------APDKLSKDYAVLAETYYQDG----TVVNYYTSFPSGTNDW 1464

Qy  731  -------DKVHLSMSKSYDNWGQIANTHWSYGVSEKITVDPITLTATKQLQSNS 777
Db 1465  NRSAAVIPAKKPIQKIEIFLLFRKNNKG----------KVWFDDIRLLEGNALIKNE 1511

Qy  778  NNVQTGKEVTTYTPSQQPIQITLFDERAGHLQSCHTLTRDGWDRVRKETDAIGQCTIYQD 837
Db 1512  YD-NDGNVVATY------------DEEGQK--NTFTYDASGNKKSETDEKGNTKLYDYN 1555

Qy  838  NYNRVIQITLPDGTIVNRKYAPFSTDTLITDIRVNGISLGGQQTFDGLSRLTQSQDGGRVW 897
Db 1556  KDNLLTKVTLKNGTSVNYRYD- 1576

Qy  898  AYTYSAGNDQCPSTVITPDGQFIHYQYQPELDDAVLQVASNEITQQFSYNPVTGALLKAV 957
Db 1577  ---HNGNTTEKSVMFGGKTQTHKYEYDVDNKNTVYIDALNRRIEN-TYDENANKIKTKM 1631

Qy  958  AEGQSLTPIYYPSGRLKMENINDMKKMSYLWTLRGLENGYTDLTGTIQKISRDTHGRVTQ 1017
Db 1632  PNGSILESVYDTADRVVGEK----------RNGKDSFT----FERDQNGQVTK 1670

Qy 1018  IKD--SSIKTLNYDDLNRHIGSQVTDLATGHMLTTVEFDGLNRBIGRKLCDSSGHTLD 1075
Db 1671  VKDLVNGVERTKTYDKADR----VTS-AT------------DSRGGKID 1702

Qy 1076  IQQSWL----------KTQQLANRIVKLNGVLQRTEQYSYDSRNRLNQ-YKCDGAECPTDKY 1126
Db 1703  ----WAYHDKANSKTEKLKEQVTQGGY---TNKVSYD-YNTLDQNIRVTDGSQ- 1748

Qy 1127  GHSIVTQNFTYDIYGNITACHTTFADGTEDHATFKF--ANP---TDPCQLTEVHHT 1177
Db 1749  -----TYRFDYDDQGNV----RTYTAGNGSGSTFNYDQANKIKDLVVGTSNSILLSERYE 1799

Qy 1178  HPDMPDNIRLKYDKAG-------------RVINITDNHGNTENFTYDTLGR------LQ 1217
Db 1800  YDQSGNRTKIKHEGAGGKVTENFVVDPINQLLNEVLPNGTTKSYTYDGFGNRTSVKVIE 1859

Qy 1218  NGQGS----VVGYGDPLNRLVSQKTDTLDCEL- 1244
Db 1860  NGKETKSIAATFNEGNQLVKFGNESLLTYDVNGNRTSDGKYKYTWNEDDQIVAITKQGENN 1919

Qy 1245  ------YYRETMLVNEVRNGEMIRLLRTGETI 1274
Db 1920  AFATYKYDEDNRRIEKNVNGQVTRYFYDGDSINPLYETDGNGTVLRQYVYSADGARLAMK 1979

Qy 1275  RASKVLLTGTDSQQSVI--LTSDKQNLSQEAYSAYGK------HKSTANDASILGYNGERA 1327
Db 1980  AQGQTLYYHNPRGDVVAMTNQDKEVVATYEYDAWGNVLTSDTKGIAAD-NPFGYAGYMY 2038

Qy 1328  DPVSGVTHLGNGYRSYDPTLMRFHTPD------SLSPPGAG----GINPYSYCLGDPINR 1377
Db 2039  DKEIGMYYL-----IARYYNPEHGVFLSVDPPDGDEDDPVTMNGYYTADNNPVMM 2088

Qy 1378  SDPSGHLSWQAWTGIGMGIAGLLLTIATGGMAIAAAGGIAAAIASTSTTALAFGALSVTS 1437
```

```
Db  2089  TDPDGKWAM------------------------LVPVVIAGAMVAARFGAKYAI----RYGAKYGKKAVKS  2131
             :||||..|           |::              :||      ::|     |:|
Qy  1438  DITSIVSGALEDASPKASSILGWVSMGMGAAGLAESAIKGGTKLLATHLGAFAEDGENALL  1497

Db  2132  ------------GWDY----GKKVAKSGWNKGKSIAQKIPRIHKVGR---I  2163
                      ||       :||:
Qy  1498  KSTSESSRIKWGVTRSLDREIVRNEEGQVIKDHSRGYTDNFMGKGEQAILVHGDKDGFLY  1557
                                              ||
Db  2164  ------------------------SFEFHTPHNGHGY  2203
               |.  .:|:
Qy  1558  HTEGNKHNG-KGPYTR  1572
          ||   :|:||
Db  2204  HLQKNKYSKYQGKWNR  2219
```

```
RESULT 8
US-10-418-861B-55
; Sequence 55, Application US/10418861B
; Publication No. US20040010131A1
; GENERAL INFORMATION:
; APPLICANT: da Silva, Ana Claudia Rasera
; APPLICANT: Farah, Shaker Chuck
; APPLICANT: Quaggio, Ronaldo Bento
; APPLICANT: Reinach, Fernando de Castro
; APPLICANT: Ferro, Jesus Aparecido
; APPLICANT: De Oliveiro, Julio Cezar Franco
; APPLICANT: De Laia, Marcelo Luiz
; APPLICANT: Setubal, Joao C.
; APPLICANT: Furlan, Luiz Roberto
; TITLE OF INVENTION: Isolated Xanthomonas nucleic acid molecules, proteins encoded the
; TITLE OF INVENTION: uses thereof
; FILE REFERENCE: FAPSP 205.1 US
; CURRENT APPLICATION NUMBER: US/10/418,861B
; CURRENT FILING DATE: 2003-04-17
; PRIOR APPLICATION NUMBER: US 60/374,620
; PRIOR FILING DATE: 2002-04-22
; NUMBER OF SEQ ID NOS: 85
; SEQ ID NO 55
; LENGTH: 1510
; TYPE: PRT
; ORGANISM: Xanthomonas
; FEATURE:
US-10-418-861B-55

  Query Match             4.1%;  Score 361;  DB 15;  Length 1510;
  Best Local Similarity   21.2%;  Pred. No. 1.3e-15;
  Matches  304;  Conservative 158;  Mismatches 466;  Indels 504;  Gaps 75;
```

```
Qy   144  VLTGFNNNAFDLK------------VPKKLLNPAGHAIYIDWNFEATQPRLNRIYDDLDG  191
               :|.            |                |..||           |||
Db   250  ILTRPNGNFYTFKKSANGTWTPDWDVRETLSEVQDNGTLTGWQVTATDDSKEQF--DLLDG  307

Qy   192  HDIPLLNLEYQGLIKTILTLFPGQKEGYRTELRFLNRQLNSIHN---------------F  236
          ||                         ::|        :|:
Db   308  ---KLTGISY-----------TDGEQLTLTYAGRQLQSVTDTRGRRLLFAYQADRIT  350

Qy   237  SLGNENPLTWSFGYTPIGKNGILGQWITSMTAPGGLKETV---NYSNNNQGHHFPQS---  290
                  ::.||         |:|       |                   ||
Db   351  QVGLPDGNVLAYGYDSQARL----RSVTLQTAARGAVASAIAGYDYGN---ARFPDALTT  402

Qy   291  ---------ANLPVLPYVTLMKQVPGAGQPAI-QAEYSY----TSHNYVGG-----GSN  330
                    :|:|   ||:      ::|          :|..||       |:|
Db   403  HRDEQDQVYASWTYDAQQRVRSVHGDPTGKIDEATIAYSGNTISVSNALGNAVTRTGIS  462

Qy   331  GIWNNKLDNLYGLMTEYNYGSTRSRR-----YKDKEGHDQIVRIERTYNNYHLLTSECKQ  385
          ||:                              .:||     |:           :|:
Db   463  KLGQAKVTAVQGLCPECTVGAFKSRSYDGNGYPDQEVDFAGVATDWRYNNRGLLASKIEA  522

Qy   386  QN---GYIQTTETAYYAIIGHNFDSQPSQFQLPKTKETWRSAD-----------NSYRSE  432
          ::         :|:|:||||        ::|            |::||      :||
Db   523  DNSSGGQKRTLQT------------DWHPS-FRVP-TDQRYDASDILVARTSWTYNSRGQA  570
```

```
Qy   433  ITETTFDESGNPLTKVIKDKKTQKIISPSTHWEYYPPAGEVDNCPPEPYGFTRFVKKIIQ  492
                :.  ||              ||::                 |  :|:
Db   571  LTVSRTDPSGSGP-TRV----TTQRYCEDSD----TAAG--NC-  601

Qy   493  TPYDSEFKDPDPEKFIQYRYSLIGSQSHVTLKIEERHYSATQLLNSTLFQYNTDKSELGRL  552
                  |:|             :||.||::      .||
Db   602  -------------SLPG----------LLLASDGARTDIADSTSYTY------YL  627

Qy   553  LKQTECT-------------KGENGKTYSVVHKFT-YTKQDD----TLQQSHSITTHDNFTIH  597
          |:|:              ||  |||:||:      ||||       :|::|  :|||||  |
Db   628  ADDASCTASASTCPHRKGDRWKVTNALGQTTEYLAYDGAGRPLSIKDTNGIVT--DYTYH  685

Qy   598  ------RSQVRSRYTGRLFSDTDTKDIVTQMSYDKLGRLLTRTLNSGTPYANTLTYDYEL  651
                :|||             |::  :|||     :|
Db   686  PRGWLTASKVRG------ADASSEADDRITRIDYWPTG-LVRQVTQPDGAFTAFTYD-AA  737

Qy   652  NNLQDDNRPPFVITTTDVNGNQLRNEFDGAGRHVSQCLKDSDGDGKFYTIHTQQYDEQGR  711
                      :|:.               ||:|              ::.|:
Db   738  HRLTD-------ITDNAGNTVHYTLDNAGNRVKEDTKDAAATLK--RTLSRVYNQLGQ  786

Qy   712  ANTHWS  753
              :|:
Db   787  LKTQATAASDPTDFAYDANGNATKVTDALATATQSEYDPLNRLSHTLQDVAGIKADTKFA  846

Qy   754  YGVSEKIT--VDPITLTATKQLQSNSNNVQ-----TGKEVTTYTPSQQPIQITLFDEAGH  806
               ::|                           ||  :|:
Db   847  YDALDNLTKVTDPKGLDTTYDYNGFGELVKLTSPDTG--VTSYT-----YDSAGN  894

Qy   807  LQSCHTLTRDGWDRVRKETDAIGQCTIYQDNYNRVIQITLPDGTIVNRKYAPFSTDTLI  866
            :                                               |||  :.
Db   895  RAT------QTDARGNTTAYSYDALNRLTKVTYPTS-------SLDVTY  930

Qy   867  T-DIRVNGISLGQQTFDGLSRLTQSODGGRVWAYTYSAGNDQCPSTVITPDGQFIHYQYQ  925
          |         :||   |  |:||             :||||              :|::
Db   931  TYDVTQTACTSG-ETF-SIGRLTKMQDGAITQYCYNRFGD-LVRKVQTSNGTALVLRYD  987

Qy   926  PEL-----------DDAVLQVASNE--ITQQFSYNPVTGALLKAVAEGQSLTPIYYPSGRL  973
            :                        :|                   ||
Db   988  YTVGGQLRMTYPDGAVVDYVRNAQGQTTQVGVTPAGGS--RQVLLGNA--TYYPFGPA  1042

Qy   974  KMENINDMKKMSYLWTL------------RGLENGYTDLTGTIQKISRDTHGRVTQIKDS  1021
                                        :|||
Db  1043  AGWTYGNGRTLARQYDLDYRPQAIQDTRPCGLDVGF-------GFPDAGNLJTALTPA  1092

Qy  1022  SIKLT----LNYDDLNRHIGSQVTDLATGEMLTTTVBFDGLNREIGRKLCDSSGHTLDIQQ  1078
           |::         :|        |                  |:
Db  1093  GNTTPEIGLGYDALGRLTG--LTDGVTGVI---------DG--  1123

Qy  1079  SWLKTQQLANRIVKLNGVLQRTEQSYDSR-NRLNQYKCDGAECPTDKYGHSIVTQNFTY  1137
                             |:||         :||        ||
Db  1124  --------------------------YSYDATGNRL-----STKVG--TATQIYTY  1146

Qy  1138  DIYGNITACHTTFADGTEDHATFKFANPTDPCQLTEVHHTHPDMPDNIRLKYDKAGRVIN  1197
                                             |:|        |:
Db  1147  --------------------------PNDSHRLSAV------AGVARIYDATG--N  1168

Qy  1198  ITDNHGNTENFTYDTLGRLQNGQ-GSVGYDPLNRLVSQKTDTLDCELYYRETMLVNEV  1255
                                       :|:|               |
Db  1169  TTAIGGTARQYTYDTSGRMTQARRAGAV------------TNNYRYN  1203

Qy  1256  RNGEMI-RLLRTGETIIAQORASKVL----LTGTDSQQSVIL---  1292
             :|::                           |:|||::||
Db  1204  GKGEQVRRFLGTTNTYTLFDEAGHWLGDVDSNGAPKQQAIWLDDLPVGLLANANKLHYIE  1263

Qy  1293  ---------TSD----KQNLSQEAYSAYGHKSTAND--ASILG--YNGERADPVS  1331
                                          ::|               :
Db  1264  PDHLGSPRVVIDPTRDVAVWTWSLKGEAFGNTAPNQDPDCDGAAAVLDMRFPGQRFDAAS  1323

Qy  1332  GVTHLGNGYRSYDPTLMRFHTPDSLSPFG-AGGINPYSYCLGDPINRSDPSG  1382
          ||:|||||          :.:|                 ::|:
Db  1324  GLNQ--NYPRDYEAATGRYGQSD--PIGLEGGISTYSYSLSSPVKYIDVLG  1370
```

RESULT 9
US-10-282-122A-76214
; Sequence 76214, Application US/10282122A
; Publication No. US2004029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 76214
; LENGTH: 843
; TYPE: PRT
; ORGANISM: Salmonella typhi
US-10-282-122A-76214

Query Match   4.0%; Score 353.5; DB 15; Length 843;
Best Local Similarity 22.8%; Pred. No. 1.7e-15;
Matches 223; Conservative 121; Mismatches 308; Indels 327; Gaps 48;

Qy 542 YNTDKSELGR----LLKQTECTKGENGKTYSVVHKFTYTKQDDTLQQSHSITTHDNFTIH 597
            ::|| | | :: :| :   :
Db   32 HHTDSGEQYRLDWNLAERSLCVTDSMGRT------RCHWWDAQGLVTAY 74

Qy 598 RSQVRSRYTGR-------LFSDTDTKDIVTQMSYDKLGRLLTRTLNSGTPYANTLTYDYE 650
            | | | | | :   :| | :|| |||
Db   75 RDEAGQMTTFRWSDEERLLLGMTDAQGGKWRYVYDRLGHL----------TETHD-P 120

Qy 651 LNNLQDDNRPPF---VITTTDVNGNQLRNEFD----------GAGRHVSQCL 689
            |:   ::   :|| ||
Db 121 LGRVEQTQWHPVWHQPETEVDAAGVAWRYEYDERGNLQAVSDPLHQRTVYGYDRH-GQVV 179

Qy 690 K--DSDGDGKFYTIHTQQVDEOG----RH------HTSTYSDYLTNGRQQTDPDKVHLSMS 738
            | |:| : | :| | ::| : | ||
Db 180 RITDARGGDKYL-----QWNEDGQLMRHTDCSGSQTAWFYDERTRLERVTDAES--NSTR 232

Qy 739 KSYDNWGQIANTHWSYGVSEKITVDPITLTATKQLQSNSNNVQTGKEVTTYTPSQQPIQI 798
            :||| | | | : ||
Db 233 YSYDGNGHLTEVMFADGRTERYQPD-----------------AAGRLVKYTSPAG---QI 272

Qy 799 TLFDEAGHLQSCHTLTRDGWDRVRKETDAIGQCTIYQYDNYNRVIQITLPDGCTIVNRKYA 858
            | : || |: : :   :   : |
Db 273 TRWQ-------RDGQGRVRRQTDATGRRTAVEYDAYGRLTTLTNENGESYRFRY- 319

Qy 859 PFSTDTLITDIRVNGISLGGQTFDGLSRLTQSQD-GRVWAYTYSAGNDQCP------- 909
                    | : | : :|| :| :||:||: | :::
Db 320 ------------DVLDRVTEQTDPGGSRRAYGYNALNAVTAVIYGGERG 356

Qy 910 ------------STVITPDGQFIHYQYQPELDDAVLQV--------ASNEITQQF 944
                         :|| || | | : || :|: |||:||||
Db 357 GEIRHGLERDAAGRLTAKITPE---TRTEYRYDAADRLLEIRRRRHDAAEGGEPEVI-RF 412

Qy 945 SYNPVTGALLKAVABG---------QSLTPIYYPSGRLKMENINDMKKMSYLWTLRGLEN 995
            |: :|| || | ||| | | |:
Db 413 SYDSAGNLLSEETAQGVLQHRYDVQGNRTETQMPDGR------------TLRYLYY 456

Qy 996 GYTDLTGTIQKISRDTHGRVTQIKDSSIKTLNYDLNRHIGSQVTDLATGMLTTTVEF 1055
            | | | :   ::   :   :
Db 457 G----SGHLQQI------------NLGRDVISEFTR-------- 476

Qy 1056 DGLNREIGRKLCDSSGHTLDIQQSWLKTQQLANRIV--KLNGVLQRT--EQYSYDSRNR 1110
            | ||:   : | ::   :: | ::| |:   :|:||:
Db 477 DHLHREVQR------SQGR-LDMRRMYDRTGRLTRKLTCKGMRGVVPETFIDRBYAYSGQDE 531

Qy 1111 LNQYKCDGAECPTDKYGHSIVTQNFTYDIYGNITAC-HTTFADGTEDHATPKFANPTDPC 1169
            | ||: || || | |: || ||| |||:   : | | |
Db 532 LLK----------KRHSRQGTDYF-YDTTGRITACRNEAYLDSWQYDAA--ANLLDRR 577

Qy 1170 Q-----------LTEVHHTHPDMPDNIRLKYDKAGRVINITDNHGNTENFTYDTL 1213
            | ||| :   :: :|||:|||::| || ::: | :
Db 578 QGETAQAGAGSVVPFNRITSYRGLH----------YRYDEYGRVVEKRGRNG-TQHYRWDAE 628

Qy 1214 GRL-----QNGQGSVYG--YDPLNRLVSQKTDTLDCBLYRETMLVNEVRNGEMIRLLR 1265
            .|| | :|: | || :|| || | || :   ::| |||::|| ||
Db 629 HRLTEVAVTRGGTVRRYGYVYDAPGRRVEKEHELDAEGKPYNRTTFLWDGMRLAQECRLGR 688

Qy 1266 TGBTIIAQQR-----------ASKVLLTGTD-SQQSVILTSDKQNLSQEA-YSAY 1307
            | |: | :| |: |: | | ||| |:|| || :| | :|
Db 689 SSSLYIYSDRGSHEPLARVDRAAPGEADEVLYYHTDVNGAPEEMTDGGGNIVWEAGYQVW 748

Qy 1308 GK---HKSTANDASIGYNGERADPVSGVTHLGNGYRSYDPTLMRFHTPDSLSPFG-AGG 1363
            || |:   :: :| | | | | || ||||::|| | |||| ||||
Db 749 GNLTHEKETRPVQQNLRFQGQYLDRETGLHY--NLYRFYDPDIGKPFISGD---PIGLAGG 803

Qy 1364 INPYSYCLGDPINRSDPSG 1382
            || |:| :|| | ||| |
Db 804 INLYQYA-PNPLSYIDPLG 821

RESULT 10
US-10-282-122A-47600
; Sequence 47600, Application US/10282122A
; Publication No. US2004029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21

```
Db     994  ---------DVSGALTQVG----QLTIRRDGLGREIGREAGDFVAQQYDALGRIRRQIAG 1041
                                 :  :        ::
Qy    1051  TTVEFDGLNREIGRKLCDSSGHTLDIQQSWLKTQQLANRIVKLNGVLQRTEQYSYDSRNR 1110
                                    :: ||     :||
Db    1042  PAVAFDALQADPARAL------------------------EQLTRQV------------YHYDAAGQ 1072
            :  :      :  :                              ||  :
Qy    1111  LNQYKCDGAECPT---DKYGHSIV------TQNFTYDIYGNITACHTTFADGTEDHATF 1160
               :: |::   |  ||                  :::||||||||
Db    1073  LERVD-TGADTLTYQRDERGQICAESLLQPSEHFRYDAVMNI-AAHGQRA--------- 1121
               ||    |  |:         :   ||| :||||
Qy    1161  KFANPTDPCQLTEVHHTH---PDMPDNIRLKYDKAGRVINIT-DNHG---NTENFTYDT 1212
               |:          ::        :|||  |     :  :||  : :::
Db    1122  ----PVD------AHHYRRGGLPEQVGYARYKYDARGRTIEKTVBQPGVRPKTWQYTWDG 1171
                            :  :          | :          |||:|
Qy    1213  LGRL------QNGQGSVYGYDPLNRLVSQKTDTLDCELYYRET------MLVN---EV 1255
            ||||         :  :          :|||          |||
Db    1172  LNRLVRVVTPERGVWA-YRYDAFNRRIEKQ-----QVGGRETVKFLWDGPMLAERWIEQ 1224
               |||             : :::  |                 ::  |
Qy    1256  RNGEMIRLLRTGETI----------IAQQRASKV--LLTGTDSQQSVILTSDKQNLSQE 1302
               :||                    :::||          ||||:
Db    1225  RDG------TTGQVVTWHIEPGSFLPLAQETDDGLFPIITDQIGRPKTVFDEQGRPVWKA 1278
               :          :||           :|    :||||
Qy    1303  AYSAYGK--HKSTANDA----SILGYNGERADPVSGVTHLGNGYRSYDPTLMRFH 1351
               :||:     :||||       ||||:           :    ::
Db    1279  AYSLWGKLLPVKRPANDADCGATSIDTTLRFSGQWADDETGLNYNLNRY--YDPDSGQYL 1336
                :||| :      ||||            :   ||||||:          ::  |||

Qy    1352  TPDSLSPFG-AGGINPYSYCLGDPINRSDPSG 1382
            :  :|||     ::|||||||
Db    1337  SAD---PIGLLGGARTQAY-VHDPSQWIDPLG 1364
```

```
Query Match             3.9%;   Score 350;  DB 15;   Length 1515;
Best Local Similarity   22.1%;  Pred. No. 7.6e-15;
Matches  232;  Conservative 126;  Mismatches 314;  Indels 380;  Gaps   59;

Qy    481  YGFTRFVKKIIQTPYDSEFKDDPEKFIQYRYSLIGSQSHVTLKIEERHYSATQLLNSTLF 540
                                                        :  |:
Db    543  FQFCQFYYAYDEHGWMTQWRDTDQTDVRYRYDTAG------RVVE--------TGTRQ 586
            ::   |                                      ::|
Qy    541  QYNTDKSELGRLLKQTECTKGENGKTYSVVHKFTYTKQDDTLQQSHS----ITTHDNFTI 596
            ::|       |   ||                            ::   :|||
Db    587  GYHT----GRFIYEAGCT-----RVIDVDGEWTYAYNDEGLVTAETDPLGHCTYSEWEL 636
                  ||:                     |      ::
Qy    597  HRSQVR--------SRY-TGRLFS--------DTDTKDIVT---------QMSY 624
               |:         |:  ||||         |:||:|             :::
Db    637  GRLMARIDPLGRRTDYRYDERGQLTSVVESSGRTVDFDYDDEQRLTGARLPNGGTIKLEY 696
               |  ||    |  |                |   |:|
Qy    625  DKLGRLLTRTLNSGTPYANTLYDY----EL---------NNLQDDNRPPFVITTTDV-- 669
               ||| |   ||             ||            |:||      |||
Db    697  DHLJSRLIART----EPDGNKTTYRYGPRGELLRVVQGDRETRLDYDDR----LRLITDIEL 748
               :|||                  :|  |:                    |||:|
Qy    670  -NGNQLRNEFDGAGRHVSQCLKDSDGDGKFYTHTQQYDEQGRHHTSYSDYLTNGR--- 725
               ||                 |:|||            |  ::|:  ::||
Db    749  PTGARFRRKIDALGR----LLEETSPDG---HVTRYD--YADGPANPRGLL 790
            |||               ::                    ||
Qy    726  -QQTDPDKVHLSMSKSYDNWGQIANTHWSYGVSEKITV--DPITLITATKQLQSNSNNVQ 781
               ::     ::|             |   :|              |:|
Db    791  SAVTRPD---------GSVSRARYN--SESLPVEWIDPLGRTIQR------ 824
            ::|                    :|      :|||||||
Qy    782  TGKEVTYTPSQQPIQITLFDEAGH---------LQSCHTLTRDGWDRVRK 823
            ||                      |||        ||||||||||
Db    825  ------TYGPFD--LLTASIDAAGHATRPEYDHATRLTKVINALGETYTYRYDAAGRLAA 876
                  :|||     |||||            ||||         ::|||
Qy    824  ETDAIGQCTIYQDYNYNRVIQITLPDGTIVNRKYAPFSTDTLITDIRVNGISLGQ--QTF 881
               |  |     |            |||||                     |    ||
Db    877  EIDWGGRATEYDRDAVGRLLLTKTLPDG--------GQWRYTY 910
            ||                      |||||           ||:|:|
Qy    882  DGLSRLTQSQDGGRVWAYTYSAGNDQCPSTVITPDGQFIHYQYQPELDDAVLQVASNEIT 941
               |  |             |  :            ||
Db    911  DASDRLIEIDAGDVKLAYRYDASG----------RLASAEVQGEHTHVT 949
               ::|||                     :   ||||||||||:|
Qy    942  QQFSYNPVTGALLKAVAEGQSLTPIYYPSGRLKMENINDMKKMSYLWTLRGLENGYTDLT 1001
               :|              ||||          |        ||||||||:
Db    950  -RFAYDR-NGRLIGEDQHGELLRHVYDADGQRRL-RMTPRRETTYAY----------- 993
               :|||                       |||           ::
Qy   1002  GTIQKISRDTHGRVTQKDSSIKTTLNYDDLNRHIGSQVTDL--------ATG----HMLT 1050
               :  |: |  || |     ||:||   ||  ||               |   ||:
```

```
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
;  SEQ ID NO 43059
;  LENGTH: 1397
;  TYPE: PRT
;  ORGANISM: Escherichia coli
US-10-282-122A-43059

Query Match     3.9%;   Score 342;  DB 15;  Length 1397;
Best Local Similarity   21.4%;  Pred. No. 2.4e-14;
Matches  256;  Conservative 157;  Mismatches  382;  Indels 404;  Gaps  63;

Qy    418 KTETW-RSADNSYRSEITETTFDESGNPLTKVI-----------KDKKTQKIIS----PS----- 461
              ::::: ||| ||  ::  |      ||||   ::           ::  :::
Db    332 QVRSFPTYDDKYRGRMVAHRHTGRPE--ICVRYD---SDGRVTEQLNPAGLSYT-------- 379

Db    218 RTQTFHREAAGEFSGEITGVT-DGAGRHFRLVLTTQAQRAEEARQQAISGGTEPSAFPDT 276
                                                                      :::
Qy    540 FQYNTDKSELGRLLKQTEC--TKGENGKTYSVVHKFTYTKQDDTLQQSHSITTHDNFTIH 597
                                                         :   ||:||||:
Db    380 YQYEKDRITITDSLNRREVLHTQGEGG-LKRVVKK----------------EHADGSVT 421

Qy    462 ----------------THWEYPPAGEVDNCPPEP----YGFT-------RFVK 488
                            : ||   ::         |       |
Db    422 QSQFDA--VGRLRAQTDAAGRTTEYSPDVVTGLITRIT---TPDGRASAFYY------- 468

Qy    277 LPGYTEYGRDNGIRLLSAVWLTHDPEYP-----ENLPAAPLVRYGWTPRGELAAVTDRSNT 331
              |||   :|    : ||   : |     |      ||  |  | | :|:  :| :|
Db    489 KIIQTPYDSEFK----------DDPEKFIQYRYSLIGSQSHVTLKIEBERHYSATQLLNSTL 539

Qy    713 HTSTSYDLTNGRQQTDPDKVHLSMSKSYDNWGQIANTHWSYGVSEKITVDPI-TLTATK 771
              :  |: |  |     |  |    :  : | | |   |  | |    :| :|   |:|:
Db    529 KTMTWSRY---GQLLSFTDCSGYVTRYDHDRFGQVTAVHREEGLSQYRAYDSRGQLIAVK 585

Qy    772 QLQSNSNNVQTGKEVTTYTPSQQPIQITLFDEAGHLQSCHTLTRDGWDRVRKETDAIGQ- 830
              |       | |    ||    :|    |         |:   |  | |    |   |
Db    586 DTQGHETRYE-------------------YNAAGDLTT--VIAPDG-SRNGTQYDAWGKA 623

Qy    831 -CTI-------YQYDNYNRVIQITLPDGTIVNRKYAPFSTDTLITDIRVNGISLGQQTF- 881
                ||        |: ||| | ||| :|   |    :|  | | | :|  |   ||||
Db    624 ICTTQGGLTRSMEYDAAGRVIRLTSENGS----HTTFRYDVLDRLIQETGFDGRTQRYH 678

Qy    882 -DGLSRLTQSQDGGRVWAYTYSAGNDQCPSTVITPDGQFIHYQYOPELDDAVLQVASNEI 940
                |:   |:| :     :  |:|| :      |  ::    |  | | :|  :|  :
Db    679 HDLTGKLIRSEDEGLV---------------------THWHYD-EADRLTHRTVNGET 714

Qy    941 TQQFSYNPVTGAL--LKAVAEGQSLTPIY-YPS-GRLKMENI--------NDM---KKMSY 986
              | |  || | |   ||     | |   |   |   |  ||:       ||     |: |
Db    715 AERWQYDE-RGWLTIDSHISEGHRVTVHYGYDSKGRLASEHLTVHHPQTNELLMQHETRH 773

Qy    987 LWTLRGLENGYTDLTGTIQKISRDTHGRVTQIKDSSIKTLNYDDLNRHIGSQVTDLATG 1046
              | ||:|  ||  | |   :|    | ||| |:| : |:| : | |:|    |  |  :|
Db    774 AYNAQGLAN--RCIPDSLPAVEWLTYG---------------------SGWLSGMKLG 808

Qy    1047 HMLTTVEF--DGLNREIGRKLCDSSGHTLDIQQSWLKTQQLANRIVKLNGVLQRTEQYS 1104
               | ||  |    |:| ||  | | :  | || : |  :| : :||:  |    |  ::
Db    809 D--TPLVEYTRDRLHRETLR----SFGRYELTTAYTPAGQLQSQ--HLNSLLS-DRDYT 858

Qy    1105 YDSRNRLNQYKCDGAECPTDKYGHSIVTQNFTYDIYGNITACHTTFAD------GTEDHA 1158
                | |||  :     |  |||: | :|    ||  || | : | | |        ||
Db    859 WNDNGELIR------ISSPRQ------TRSYSYSTGRLTGVHTTAANLDIRIPYTTDPA 906

Qy    1159 TFKFANPTDPCQLTEVHHTHPD-----MPDN-------IRLKYDKAGRVINITD----- 1200
               |  |  | | | | :|  ||                 :| |  |  ||
Db    907 GNRLLPDP---------ELHPDSALSMWPDNRIARDAHYLRYDRHGRLTEKTDLIPEGV 956

Qy    1201 ---NHGNTENFTYDTLGRLQNGQGSVYG---------YDPLNRLVSQKTDTLDCEL---- 1244
```

```
Db    957 IRTDDERTHRYHYDSQHRLVHTVHTRTQYAEPLVESRYLYDPLGRRVAKRVWRRERDLTGWM 1016
              :||:      :   ::  |:  || ||   :|         |  :     :  |   ::|
Qy   1245 ------------YY---RETMLVNE------VRNGEMIRLLR----TGETIIAQQRA---- 1276

Db   1017 SLSRKPQVTWYGWDGDRLTTIQNDRTRIQTIYQPGSFTPLIRVETATGELAKTQRRSLAD 1076
              :| ||| ||| |   ||: | | | || |     ::|  | ||| || || |||::|
Qy   1277 --------------SKVLLTGTDSQQSVILTSDKQNLSQEA----- 1303

Db   1077 TLQQSGGEDGGSVVFPPVLVQMLDRLESEILA---DRVSEESRRWLASCGLTVAQMQSQM 1133
              || :   | |  |  | :|  :| |  | |:|     | | |   |:  |  |||:|
Qy   1304 ----------------------------------YSAYGKHKSTANDAS-----ILGYNG 1324

Db   1134 DPVYTPARKIHLYHCDHRGLPLALISTEGTTAWYAEYDEWGNLLNEENPHQLQQLIRLPG 1193
              | :|| |  | |    ||||  | ||:|  ||  | : :| | | |    ||:| |||
Qy   1325 ERADPVSGVTHLGNGYRSYDPTLMRFHTPDSLSPFG-AGGINPYSYCLGDPINRSDPSG 1382

Db   1194 QQYDEESGLYY--NRHRYDPLQGRYITQD---PIGLKGGWNFYQYPL-NPISNIDPLG 1246
```

```
RESULT 12
US-10-282-122A-43060
;  Sequence 43060, Application US/10282122A
;  Publication No. US20040029129A1
;  GENERAL INFORMATION:
;  APPLICANT: Wang, Liangsu
;  APPLICANT: Zamudio, Carlos
;  APPLICANT: Malone, Cheryl
;  APPLICANT: Haselbeck, Robert
;  APPLICANT: Ohlsen, Kari
;  APPLICANT: Zyskind, Judith
;  APPLICANT: Wall, Daniel
;  APPLICANT: Trawick, John
;  APPLICANT: Carr, Grant
;  APPLICANT: Yamamoto, Robert
;  APPLICANT: Forsyth, R.
;  APPLICANT: Xu, H.
;  TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
;  FILE REFERENCE: ELITRA.034A
;  CURRENT APPLICATION NUMBER: US/10/282,122A
;  CURRENT FILING DATE: 2003-02-20
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/230,335
;  PRIOR FILING DATE: 2000-09-06
;  PRIOR APPLICATION NUMBER: 60/230,347
;  PRIOR FILING DATE: 2000-09-09
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/267,636
;  PRIOR FILING DATE: 2001-02-09
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 78614
;  SOFTWARE: PatentIn version 3.1
;  SEQ ID NO 43060
;  LENGTH: 1411
;  TYPE: PRT
;  ORGANISM: Escherichia coli
US-10-282-122A-43060

Query Match     3.8%;   Score 340.5;  DB 15;  Length 1411;
Best Local Similarity   20.4%;  Pred. No. 3.1e-14;
Matches  287;  Conservative 172;  Mismatches  459;  Indels 490;  Gaps  70;
```

```
Qy   1304 ------------------------------------------------------YSAYGKHKSTAND--ASILGYNGERADP 1329
Db   1139 PARKIHLYHCDHRGLPLALISTEGATAWCAEYDEWGNLLNEENPHQLQLIRLPGQQYDE 1198
Qy   1330 VSGVTHLGNGYRSYDPTLMRFHTPDSLSPFG-AGGINPYSYCLGDPINRSDPSGHLSWQA 1388
Db   1199 ESGLYY--NRHRYYDPLQGRYITQD--PIGLKGGWNLYGYQL-NPISDIDPLGLSMWED 1252
Qy   1389 WTGIGMGIAGLLLTIATGGMAIAAAGGIAAAIASTSTTALAFGALSVTSDITSIVSGALE 1448
Db   1253 ------------AKSG---ACTNGLCGTLSA---MIGPDKFDSIDSTAY 1283
Qy   1449 DASPKASSILGWVSMGMGVSMGMGAAGLAESAIKGGTKLATHLGAFAEDGENALLKSTSESSRIKW 1508
Db   1284 DALNKINS---------------------------QSICEDKEFA--------- 1301
Qy   1509 GVTRSLDREIVRNEEGQVIKDHSRGYTDNFMGKGEBQ------AILVHGDKDGFLYHTEG 1561
Db   1302 ------------------GLICKDNSGRYFSTAPNRGERKGSYPFNSPCPNGTEKVSAYHTHG 1346
Qy   1562 NKHNGK--GPYTRHTPEQLVDYLKDNNI 1587
Db   1347 ADSHGEYWDEIFSGKDEKIVK-SKDNNI 1373
```

RESULT 13
US-09-912-020-340
; Sequence 340, Application US/09912020
; Patent No. US20020045592A1
; GENERAL INFORMATION:
;   APPLICANT: Zyskind, Judith
;   APPLICANT: Ohlsen, Kari L.
;   APPLICANT: Trawick, John
;   APPLICANT: Forsyth, R. Allyn
;   APPLICANT: Froelich, Jamie M.
;   APPLICANT: Carr, Grant J.
;   APPLICANT: Yamamoto, Robert T.
;   APPLICANT: Xu, H. Howard
;   TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
;   TITLE OF INVENTION: ESCHERICHIA COLI
;   FILE REFERENCE: ELITRA.001DV1
;   CURRENT APPLICATION NUMBER: US/09/912,020
;   CURRENT FILING DATE: 2001-07-23
;   PRIOR APPLICATION NUMBER: 09/492,709
;   PRIOR FILING DATE: 2000-01-27
;   PRIOR APPLICATION NUMBER: 60/117,405
;   PRIOR FILING DATE: 1999-01-27
;   NUMBER OF SEQ ID NOS: 485
;   SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 340
;   LENGTH: 1426
;   TYPE: PRT
;   ORGANISM: E. Coli
US-09-912-020-340

Query Match   3.8%;   Score 339;   DB 9;   Length 1426;
Best Local Similarity 20.2%;   Pred. No. 4.1e-14;
Matches 257;   Conservative 150;   Mismatches 409;   Indels 458;   Gaps 60;

```
Qy   572 KFTYTKQDDTLQQSHSITTHDNFTIHRSQVRSRYTGRLFSDTDTKDIVTQMSYDKLGRLL 631
Db   257 ::::||:|| ::    ::               ||      ::||::|
Qy   317 RYTYTBEAGELL--AVYDRSNTQVRAFTYDAQHPGRMVAHRYAGRPEMRYRYDDTGRVV 372
Db   632 TRTLNSGTPYANTLTYDYELNNLQDDNRPPFVITTTD----------------------- 668
Qy   373 EQLNPAG---------LSYRY-----LYEQDR----ITVTDSLNRREVLHTEGGAGLKRVVKKE 418
Db   669 -VNGNQLRNEFDGAGRHVSQCLKDSDGDGKFYIHTQQYD------EQGRHHTSYSDYL 721
Qy   419 LADGSVTRSGYDAAGRLTAQ--TDAAGRRTEYGLNVVSGDITDITTPDGRETKFYYND--- 474
Db   722 TNGRQQT---DPDKVHLSMSKSYDNWGQIANTHWSYGVSEKITVDPIITLATKQLQSNSN 778
```

```
Qy   418 KTETW-RSADNSYRSEITETTFPESGNPLTKVI------KDKKTQKIIS-----PS------ 461
Db   218 RTQTFHREAAGEFSGEITGVT-DGAWRHFRLVLTTQAQRAEEARQQAISGGTEPSAFPDT 276
Qy   462 -------------THWEYYPPAGEVDNCPPEP---YGFT-----RFVK 488
Db   277 ENLPAAPLVRYGWTPRGELAVVYDRSGK 331
Qy   489 KIIQTPYDSEFK-------DDPEKFIQYRYSLIGSQSHVTLKIEERHYSATQLLNSTL 539
Db   332 QVRSFTYDDKYRGRMVAHRHTGRPE--IRYRYD---SDGRVTEQLNPAGLSYT------ 379
Qy   540 FQYNTDKSELGRLLKQTEC--TKGENGKTYSVVHKFTYTKQDDTLQQSHSITTHDNFTIH 597
Db   380 YQYEKDRITITDSLDRREVLHTQGEAG-LKRVVKK--------------EHADGSVT 421
Qy   598 RSQVRSRYTGRLFSDTDTKDIVTQMSYDKLGRLLGRLLTRTLNSGTPYANTLTYDYELNNLQDD 657
Db   422 QSQFDA--VGRLRAQTDAAGRTEYSPDVVTGLITRIT---TPDGRASAFYYNHHNQ--- 473
Qy   658 NRPPFVITTTDVVNGNQLRNEFDGAGRHVSQCLKDSD-------GDGKFYIHTQQYDEQGRH 712
Db   474 ------LTSATGPDGLELRREYDELGRLLIQETAPDGDITRYRYDNPHSDLPCATEDATGSR 528
Qy   713 HTSTYSDYLTNGRQQTDPDKVHLSMSKSYDNWGQIANTHWSYGVSEKITVDPI-TLTATK 771
Db   529 KTMTWSRY---GQLLSFTDCSGYVTRYDHDRFGQMTAVHREEGLSQYRAYDSRGQLIAVK 585
Qy   772 QLQSNSNNVQTGKEVTYTYTPSQQPIQITLFDEAGHLQ----------SCHTLTRDGWDRVRKE 824
Db   586 DTQGHETRYE-------------YNIAGDLTAVIAPDGSRNGTQYDAWGKAVRT 626
Qy   825 TDAIGQCTIYQYDNYNRVIQITLPDGTIVNRKYAPFSTDTLITDIRVNGISLGQQTF--D 882
Db   627 TQG-GLTRSMEYDAAGRVIRLTSENGS-------HTTFRYDVLDRLLQETGFDGRTQRYHHD 680
Qy   883 GLSRLTQSODQGGRVWAYTYSAGNDQCPSTVITPDGQFIHYQYQPELDDAVLQVASNEITQ 942
Db   681 LTGKLIRSEDEGLV----------------------THWHYD-EADRLTHRTVKGETAE 716
Qy   943 QFSYNPVTGAL--LKAVAEGQSLTPIYY---PSGRL--KMENINDMKKMSYLM------- 988
Db   717 RWQYDE-RGWLTDISHISEGHRVA-VHYRYDEKGRLTGERQTVHHPQTEALLWQHETRHA 774
Qy   989 -TLRGLENGYTDLTGTIQKISRDTHGRVTQIKDSSIKTTLNYDDLNRHIGSQVTDLATGH 1047
Db   775 YNAQGLAN----RCIPDSLPAVEWLTYG--------SGYLAGMKLGD------------ 809
Qy   1048 MLTTTVEF--DGLNREIGRKLCDSSGHTLDIQQSWLKTQQLANRIVKLNGVLQRTEQYSY 1105
Db   810 --TPLVEYTRDRLHRETLR-----SFGRVELTTAYTPAGQLQSQ--HLNSLLS-DRDYTW 859
Qy   1106 DSRNRLNQYKCDGAECPTDKYGHSIVTQNFTYDIYGNITACHTTFPADGTEDHATFKFANP 1165
Db   860 NDNGELIR-----ISSPRQ-------TRSYSYSTTGRLTGVHTTAA----NLDIRIPYA 902
Qy   1166 TDPC--QLTEVHHTHPD-------MPDN----IRLKYDKAGRVINITD-------NH 1202
Db   903 TDPAGNRLPD-PELHPDSTLSMWPDNRIARDAHYLYRYDRHGRLTEKTDLIPEGVIRTDD 961
Qy   1203 GNTENFTYDTLGRLQNGQ--------GSVYGYDPLNRIVSQKTDTLDCEL------- 1244
Db   962 ERTHRYHYDSQHRLVHYTRTQYEEPLVESRYLYDPLGRRVAKRVWRRERDLTGWMSLSRK 1021
Qy   1245 -----YY----REIMLVNE------VRNGEMIRLLR---TGETIIAQQRA---- 1276
Db   1022 PQVTWYGWDGDRLTTIQNDRSRIQTIYQPGSFFTPLIRVETATGELAKTQRRSLADALQQS 1081
Qy   1277 -----------SKVLLTGTDSQQSVILTSDKQNLSQEA----------- 1303
Db   1082 GGEDGGSVVFPPVLVQMLDRLESEILA---DRVSEESRRWLASCGLTVBQMKNQMDPVVT 1138
```

```
RESULT 14
US-10-282-122A-42617
; Sequence 42617, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;   APPLICANT: Wang, Liangsu
;   APPLICANT: Zamudio, Carlos
;   APPLICANT: Malone, Cheryl
;   APPLICANT: Haselbeck, Robert
;   APPLICANT: Ohlsen, Kari
;   APPLICANT: Zyskind, Judith
;   APPLICANT: Wall, Daniel
;   APPLICANT: Trawick, John
;   APPLICANT: Carr, Grant
;   APPLICANT: Yamamoto, Robert
;   APPLICANT: Forsyth, R.
;   APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
;
; SEQ ID NO 42617
;   LENGTH: 1426
;   TYPE: PRT
;   ORGANISM: Escherichia coli
US-10-282-122A-42617

  Query Match       3.8%;  Score 339;  DB 15;  Length 1426;
  Best Local Similarity   20.2%;  Pred. No. 4.1e-14;
  Matches  257;  Conservative 150;  Mismatches 409;  Indels 458;  Gaps  60;
```

```
Qy      572 KFTYTKQDDTLQQSHSITTHDNFTIHRSQVRSRYTGRLFSDTDTKDIVTQMSYDKLGRLL  631
                ::|:|   :|   |  :  :  ::  :  ::  ||   |   :::|    :|    |:
Db      317 RYTYTEAGELL----AVYDRSNTQVRAFTYDAQHPGRMVAHRYAGRPEMRYRYDDTGRVV  372

Qy      632 TRTLNSGTPYANTLITYDYELNNLQDDNRPPVITTTD-------------------------  668
                 :::   |   |   ::  |     :  |
Db      373 EQLNPAG--------LSYRY-----LYEQDR------ITVTDSLNRREVLHTEGGAGLKRVVKKE  418

Qy      669 -VNGNQLRNEFDGAGRHVSQCLKDSDGDGKFYTIHTQQYD-------EQRHHTSTYSDYL  721
               :       |    |         |     |       ::
Db      419 LADGSVTRSGVDAAGRLTAQ--TDAAGRRTEYGLNVVSGDITDITTPDGRETKFYYND--  474

Qy      722 TNGRQQT----DPDKVHLSMSKSYDNWGQIANTHWSYGVSEKITVDPITLTATKQLQSNSN  778
                :  ::  :                         ::
Db      475 --GNQLTAVVSPD--GLESRREYDEPGRLV-------------------------------SE  502

Qy      779 NVQTGKEVT-TYTPSQQPIQITLFPDEAGHLQSCHTLTRDGWDRVRKETDAIGQCTTYQYD  837
```

Left column (Db/Qy pairs):

```
Db      475 --GNQLTAVVSPD--GLESRREYDEPGRLV-------------------------------SE  502
Qy      779 NVQTGKEVT-TYTPSQQPIQITLFPDEAGHLQSCHTLTRDGWDRVRKETDAIGQCTTYQYD  837
Db      503 TSRSGETVRYRYRYDDAHSELPATTTDATG---STRQMTWSRYGQLLAFTDCSGYQTRYEYD  559
Qy      838 NYNRVIQITLPDGTIVNRKYAPFSTDTLITDIRVNGISLGQQTPDGLSRLTQSQDG-GRV  896
Db      560 RFGQM-------TAVHRE----------EGISL-YRRYDNRGRLTSVKDAQGRE  595
Qy      897 WAYTYSAGNDQCPSTVITPDGQFIHYQYQ-----------------------------  925
Db      596 TRYEYNAAGDL--TAVITPDGNRSETQYDAWGKAVSTTQGGLTRSMEYDAAGRVISLTNE  653
Qy      926 --------PELDDAVLQVASNBITQQFSYNPVTGALLKAVAEGQSLTPIYYPSGRLKME  976
Db      654 NGSHSVFSYDALDRLVQQGGFDGRTQRYHYD-LTGKLTQSEDEGLVILMWYDESDRITHR  712
Qy      977 NINDMKKMSYLWTLRGLENGYTDLTGTIQKISRDTHGRVTQIKDSS-----IKTTLNYDDL  1032
Db      713 TVNGBPAEQWQY----------------------DGHGWLTDISHLSEGHRVAVHYGYDDK  751
Qy     1033 NRHIG--SQVTDLATGHML-----------------------------------------  1049
Db      752 GRLTGECQTVENPETGELLWQHETKHAYNEQGLANRVTPDSLPVEWLTYGSGYLAGMKL  811
Qy     1050 --TTTVEF--DGLNREIGRKLCDSSGH--TLDIQQSWLKTQQLANR------------  1089
Db      812 GGTPLVEYTRDRLHRETVRSFPGSMAGSNAAYELTSTYTPAGQLQSQHLNSLVYDRDYGWS  871
Qy     1090 ----IVKLNGVLQRTEQYSYDSRNRLNQYKCDGAEC-------PTDKYGHSI--------  1130
Db      872 DNGDLVRISGPRQ-TREYGYSATGRLESVRTLAPDLDIRIPYATDPAGNRLPDPELHPDS  930
Qy     1131 ----------VTQN----FTYDIYGNITACHTFADG---TEDHATFKFANPTDPCQLTEV  1174
Db      931 TLTVWPDNRIAEBDAHVYRHDEYGRLTEKTDRIPAGVIRTDDERT----------------  975
Qy     1175 HHTHFHDMPDNIRLKYDKAGRVINITD-NHGN---TENFTYDTLGR--------------  1215
Db      976 HHYH-------YDSQHRLIVFYVTRIQHGEPLVESRYLYDPLGRRMAKRVWRRERRDLTG  1025
Qy     1216 ----LQNGQGSVYGYDPLNRLRVSQKTDTLDCELYYR---------------------  1247
Db     1026 WMSLSRKPEVTWYGWDG-DRLTTVQTDTTRIQTVVEPGSFTPLIRVETENGEREKAQRRS  1084
Qy     1248 --ETMLVNEVRNG----------EMIRLL-RTGETIIA-------------QQRASKV  1279
Db     1085 LAEFTLQQEGSENGHGVVFPAEIVRLLDRLEEIRADRVSSESRAWLAQCCLTVEQLARQV  1144
Qy     1280 LLTGTDSQQS-----------VILTSDKQNLSQEA-YSAYGKHKSTANDASILGYN----  1323
Db     1145 EPEYTPARKAHLYHCDHRGLPLALISEDGNTAWSAEYDEWGNQLNEENPHHV--YQPYRL  1202
Qy     1324 -GERADPVSGVTHLGNGYRSYDPTLMRFHTPDSLSPFG-AGGINPYSYCLGDPINRSDPS  1381
Db     1203 PGQQHDEESGLYY--NRHRYYDPLQGRYITQD---PMGLKGGWNLYQYPL-NPLQQIDPM  1256
Qy     1382 GHLSMQAWTGIGMGIAGLLLTIATGGMAIAAAGGIAAAIASTSTTALAFGALSVTSDITS  1441
Db     1257 GLL--QTWDDARSG-----ACTCGV----CGVLSRIIGPSKFDSTADDAALD-------  1296
Qy     1442 IVSGALEDASPKASSILGWVSMGMGAAGLAESAIKGGTKLATHLGAFAEDGENALLKSTS  1501
Db     1297 ----ALKETQNRS-----LCNDMEYSGIVCKDTNG------KYFASKAETDNLR--K  1336
Qy     1502 ESSRIKWGVTRSLDREIVRNEEGQVIKDHSRGYTDNFMGKGEQAILVHGDK--DGFLYHT  1559
Db     1337 ESYPLKRKCPTGTDRVAAYHTHG---ADSHGDYVDEFFSSSDKNLVRSKDNNLEAFYLAT  1393
Qy     1560 EGNKH---NGKGPY  1570
Db     1394 PDGRFEALNNKGEY  1407
```

Left-most first entries:

```
Db      475 --GNQLTAVVSPD--GLESRREYDEPGRLV-------------------------------SE  502
```

```
US-09-815-242-10384
; Sequence 10384, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
;     APPLICANT:  Haselbeck, Robert
;     APPLICANT:  Ohlsen, Kari L.
;     APPLICANT:  Zyskind, Judith W.
;     APPLICANT:  Wall, Daniel
;     APPLICANT:  Trawick, John D.
;     APPLICANT:  Carr, Grant J.
;     APPLICANT:  Yamamoto, Robert T.
;     APPLICANT:  Xu, H. Howard
;     TITLE OF INVENTION: Identification of Essential Genes in
;     TITLE OF INVENTION: Prokaryotes
;     FILE REFERENCE: ELITRA.011A
;     CURRENT APPLICATION NUMBER: US/09/815,242
;     CURRENT FILING DATE: 2001-03-21
;     PRIOR APPLICATION NUMBER: 60/191,078
;     PRIOR FILING DATE: 2000-03-21
;     PRIOR APPLICATION NUMBER: 60/206,848
;     PRIOR FILING DATE: 2000-05-23
;     PRIOR APPLICATION NUMBER: 60/207,727
;     PRIOR FILING DATE: 2000-05-26
;     PRIOR APPLICATION NUMBER: 60/242,578
;     PRIOR FILING DATE: 2000-10-23
;     PRIOR APPLICATION NUMBER: 60/253,625
;     PRIOR FILING DATE: 2000-11-27
;     PRIOR APPLICATION NUMBER: 60/257,931
;     PRIOR FILING DATE: 2000-12-22
;     PRIOR APPLICATION NUMBER: 60/269,308
;     PRIOR FILING DATE: 2001-02-16
;     NUMBER OF SEQ ID NOS: 14110
;     SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10384
;     LENGTH: 1377
;     TYPE: PRT
;     ORGANISM: Escherichia coli
US-09-815-242-10384

Query Match           3.7%;     Score 331.5;    DB 9;    Length 1377;
Best Local Similarity    20.8%;     Pred. No. 1.3e-13;
Matches  288;  Conservative 170;  Mismatches 450;  Indels 477;  Gaps  75;

Qy   211 LFPGQKEGY-RTELRPLNRQLNSIHNFSLGNENPLTWSFGYTPIGKNGILGQWITSMTAP  269
         ||||  ||:  :                      |:
Db   126 LFPGE-DGYSRSESLMLVR--GGVAKLDEGHRLAALWQ----------------------  160

Qy   270 GGLKETVNYS---------------NNNQG------HHFPQSANL---PVLPYVTLMKQVPGAGQ-  310
            :    |:              :  :      :    :          ::
Db   161 -ALPEELRLSPHRYLATNSPQGPWWLLGWCERVPEADEVLPAPLPPYRVLTGLVDRFGRT  219

Qy   311 PAIQAEYSYTSHNYVGGGSNGIWNNKLDNLYGLMTEYNYGSTESRRYKDKEGHDQIVRIE  370
           :    :  : :||         | ::  |            ::|:
Db   220 QTFHREAAGEFSGEITGVTDGAWRH-----FRLVL----------TTQAQR---------  255

Qy   371 RTYNNYHLLTSECKQQNGYIQTETAYYAIIGHNFDSQPSQFQLPKT--KTETWRSADN  427
                       : ::||          :|::          : ::    : |   ::
Db   256 ----------AEEARQQ--------AISG---GTEPSAF-PDTLPGVTEYGR--DN  287

Qy   428 SYRSEITETTFDESGNPLIVKIDKKTQKIISPSTHWEYYPPAGEVDNCPPEP---YGFT  484
            | :|||     :         : :||      :||:|        ::      ||::
Db   288 GIRLSAVWLTHDPE--------------------YP----ENLPAAPLVRYGWT  317

Qy   485 ---------RFVKKIIQTPYDSEFK---------DDPEKFIQYRYSLIGSQSHVTLKIE  525
                      ::    :|||:|:|         :|| :: | :||     || :|
Db   318 PRGELAVVYDRSGKQVRSFTYDDKYRGRMVAHRHTGRPE--IRYRYD---SDGRVTEQLN  372

Qy   526 ERHYSATQLLNSTLFQYNTDKSELGRLLKQTEC--TKGENGKTYSVVHKFTYTKQDDTLQ  583
           ||:| ||  :|| ||| |:||       ||         |||: |:  :| | |::|||
Db   373 PAGLSYT--------YQYEKDRITITDSLDRREVLHTQGEAG-LKRVVKK---------  413

Qy   584 QSHSITTHDNFTIHRSQVRSRYTGRLFSDTDTKDIVTQMSYDKLGRLLTRTLNSGTPYAN  643
           ||: |  | :   :: | ::|   |||        || |:||||||||:|:     ||
```

```
Db   503 TSRSGETVRYRYRYDAHSELPATTTDATG---STRQMTWSRYGQLLAFTDCSGYQTRYEYD  559
            ::   || |:||||:|:      |:  |||    : ||| :||| |||  :   | | |
Qy
Db   838 NYNRVIQITLPDGTIVNRKYAPFSTDTLITDIRVNGISLGQQTFDGLSRLTQSODG-GRV  896
           | : :::  | ||: :::   :  : :   |:|||||:||       | | | || |  |
Db   560 RFGQM----------TAVHRE--------EGISL-YRRYDNRGRLTSVKDAQGRE  595
           |: :          |||||:        |||:| |||||||:|||| |  ::| |:
Qy   897 WAYTYSAGNDQCPSTVITPDGQFIHYQYQ------------------------  925
Db   596 TRYEYNAAGDL--TAVITPDGNRSETQYDAWGKAVSTTQGGLTRSMEYDAAGRVISLTNE  653
           | : | ||| |  |||||||| :    | | || |    |||:|  |||  |||: |  |
Qy   926 --------PELDDAVLQVASNEITQQFSYNPVTGALLKAVAEGQSLTPIYYPSGRLKME  976
                   ||: | |:|| :|  |||: | ||||| |:|   :| || | ||  |||: |
Db   654 NGHSVFSYDALDRLVQQGGFDGRTQRYHYD-LTGKLTQSEDEGLVILWYYDESDRITHR  712
Qy   977 NINDMKKMSYLWTLRGLENGYTDLTGTIQXISRDTHGRVTQIKDSS----IKTTLNYDDL  1032
            :|  :   || | |||:|||| | ||:  :||| ||||| : ||       :||:||| |
Db   713 TVNGEPAEQWQY---------------------DGHGWLTDISHLSEGHRVAVHYGYDDK  751
Qy  1033 NRHIG--SQVTDLATGHML--------------------------------  1049
           || :|  |||| ||||: |
Db   752 GRLTGECQTVENPETGELLWQHETKHAYNEQGLANRVTPDSLPPVEWLTYGSGVLAGMKL  811
Qy  1050 --TTTVEF--DGLNREIGRKLCDSSGH-TLDIQQSWLKTQQLANR-----------  1089
             |||:|:  |||:|||||||  ||:|| |||| | ||| || |||
Db   812 GGTPLVEYTRDRLHRETVRSFGSMAGSNAAYELTSTVTPAGQLQSQHLNSLVVDRDYGWS  871
Qy  1090 ----IVKLNGVLQRTEQYSYDSRNRLNQYKCDGAEC------PTDKYGHSI--------  1130
               :|||:|| | |||||:||||||||| |  | |       | | || |
Db   872 DNGDLVRISGPRQ-TREYGYSATGRLESVRTLAPDLDIRIPYATDPAGNRLPDPELHPDS  930
Qy  1131 ---------VTQN---FTYDIYGNITACHTTFADG---TEDHATFKFANPTDPCQLTEV  1174
                    |||    |||| |||||   |||| |||    |:| |:| |: | |||||:|
Db   931 TLTVWPDNRIAEDAHYVVRHDEYGRLTEKTDRIPAGVIRTDDERT----------  975
Qy  1175 HHTHPDMPDNIRLKYDKAGRVINITD-NHGN--TENFTYDILGR----------  1215
           |||| | ||||:|||||||||:|:|| |:|:  ||||||||:|||
Db   976 HHYH-------------YDSQHRLVFYTRIQHGEPLVESRYLYDPLGRRMAKRVWRRERDLTG  1025
Qy  1216 ---LQNGQGSVYGVGYDPLNRLVSQKTDTLDCELYR-----------  1247
              ||:|| |:  |||||||||| |::|||||||||
Db  1026 WMSLSRKPEVTWYGWDG-DRLTTVQTDTRIQTVYEPGSFTPLIRVETENGEREKAQRRS  1084
Qy  1248 --ETMLVNEVRNG----------EMIRLL-RTGETIIA-----------QQRASKV  1279
             |||:|:|| ||           |:|::| ||||:|||           |:|| |:
Db  1085 LAETLQQEGSENGHGVVFPAELVRILDRLEBEIRADRVSSESRAWLAQCGLTVEQLARQV  1144
Qy  1280 LLTGTDSQQS----------VILTSDKQNLSQEA--YSAYGKHKSTANDASILGYN---  1323
           | || ||| |          |||||||:| ||||  | ||||:||||||||||||||
Db  1145 EPEYTPARKAHLYHCDHRGLPLALISEDGNTAWSAEYDEWGNQLNEENPHHV--YQPYRL  1202
Qy  1324 -GERADPVSGVTHLGNGYRSYDPTLMRFHTPDSLSPFG-AGGINPYSYCLGDPINRSDPS  1381
            ||||||||| || ||||||||| ||||||||| || | |||||||| |||||||| |||
Db  1203 PGQQHDEESGLYY--NRHRYYDPLQGRYITQD--PMGLKGGWNLYQYPL-NPLQQIDPM  1256
Qy  1382 GHLSWQAWTGIGMGIAGLLLTIATGGMAIAAAGGIAAAIAST5TTALAFGALSVTSDITS  1441
           | || | |||||||||||||||||||| | ||||| |||||| |||||| |||:| |||
Db  1257 GLL--QTWDDARSG-----ACTGGV-----CGVLSRIIGPSKFDSTADAALD-------  1296
Qy  1442 IVSGALEDASPKASSILGWSMGMGAAGLAESAIKGGTKLATHLGAFAEDGENALLKSTS  1501
           |:||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1297 ----ALKETQNRS-------LCNDMEYSGIVCKDTNG---------KYFASKAETDNLR--K  1336
Qy  1502 ESSRIKWGVTRSLDREIVRNEEGQVIKDHSRGYTDNFMGKGEQAILVHGDK--DGFLYHT  1559
           || ||||| |||| || |:|||| |||| |||||||||||||||||:|||   |||||||
Db  1337 ESYPLKRKCPTGTDRVAAYHTHG---ADSHGDYVDEFFSSSDKNLVRSKDNNLEAFYLAT  1393
Qy  1560 EGNKH---NGKGPY  1570
           ||| |   |||| |
Db  1394 PDGRFEALNNKGEY  1407

RESULT 15
```

```
Db    414  ------EHADGSVTQSQFDA--VGRLRAQTDAAGRTTEYSPDVVTGLITRIT---TPDGR  462
                    :  :                       :    :: ::    : ::
Qy    644  TLTVDYELNNLQDDNRPPFVITTTDVNGNQLRNEFPGAGRHVSQCLKDSD-----GDGKF  698
                :.:           ::  :: ::: ||| | :: :                 |
Db    463  ASAFYNHHNQ--------LTSATGPDGLELRREYDELGRLLIQETAPDGDITRYRYDNPH  514

Qy    699  YTIHTQOYDEQGRHHTSTYSDYLTNGRQQTDPDKVHLSMSKSYDNWGQIANTHWSYGVSE  758
                    :       : ||::      ::   ::::
Db    515  SDLPCATEDATGSRKTMTWSRY---GQLLSFTDCSGYVTRYDHDRFGQMTAVHRREGLSQ  571

Qy    759  KITVDPI-TLTATKQLQSNSNNVQTGKEVTTYTPSQQPIQITLFDEAGHLQ-------SC  810
                ::|                  :                               :
Db    572  YRAYDSRGQLIAVKDTQGHETRYE--------------YNIAGDLTAVIAPDGSR  612

Qy    811  HTLTRDGWDRVRKETDAIGQCTIYQYDNYNRVIQITLPDGTIVNRKYAPFSTDTLITDIR  870
                : :       |  ::       |||:    :      :    :  : :
Db    613  NGTQYDAWGKAVRTTQG-GLTRSMEYDAAGRVIRLTSENGS-----HTTPRYDVLDRLIQ  666

Qy    871  VNGISLGQQTF--DGLSRLTQSQDGGRVWAYTYSAGNDQCPSTVITPDGQPIHYQYQPEL  928
                :::                      :::         :           :  :
Db    667  ETGFPDGRTQRYHHDLTGKLIRSEDEGLV--------------------THWHYD-EA  702

Qy    929  DDAVLQVASNEITQQFSYNPVTGAL--LKAVAEGQSLTPIYY---PSGRL--KMENINDM  981
                 _           :|    : :||:    : :::       |    |
Db    703  DRLTHRTVKGETAERWQYDE-RGWLTDISHISEGHRVA-VHYRYDEKGRLTGERQTVHHP  760

Qy    982  KKMSYLW--------TLRGLENGYTDLTGTIQKISRDTHGRVTQIKDSSIKTTLNYDDLN  1033
                :  ||                  :|:  :      :
Db    761  QTEALLWQHETRHAYNAQGLAN--RCIPDSLPAVEWLTYG------------SGVLAGMKLGD--  809

Qy    1034 RHIGSQVTDLATGHMLTTTVBF--DGLNREIGRKLCDSSGHTLDIQQSWLKTQQLANRIV  1091
                                    :| |::  : :||| :    : |
Db    810  ------------------TPLVEYTRDRLHRETLR-----SFGRYELTTAYTPAGQLQSQ-  846

Qy    1092 KLNGVLQRTEQYSYDSRNRLNQYKCDGAECPTDKYGHSIVTQNFTYDIYGNITACHTTFA  1151
                ||:|:            :          | ::   ||| :|||:
Db    847  HLNSLLS-DRDYTWNDNGELIR------ISSPRQ------TRSYSYSTTGRLTGVHTTAA  893

Qy    1152 DGTEDHATFKFANPTDPC--QLTEVHHTHPD------MPDN------IRLKYDKAGRVIN  1197
                ||               :: :       |||         :: |: |:
Db    894  ----NLDIRIPYATDPAGNRLPD-PELHPDSTLSMWPDNRIARDAHYLYRYDRHGRLTE  947

Qy    1198 ITD--------NHGNTENFTYDTLGRLQNGQ--------GSVVGYDPLNRLLVSQKTDT  1239
                ::            : :    :    : ::: | ::: |:
Db    948  KTDLIPEGVIRTDDERTHRYHYDSQHRLVHYTRTQYEEPLVESRYLYDPLGRRVAKRVWR  1007

Qy    1240 LDCEL----------------YY----RETMLVNE-------VRNGEMIRLLR-----TGETI  1270
                ||:                   :  : :|:         :        |||:
Db    1008 RERDLTGWMSLSRKPQVTWGWDGDRLTTIQNDRTRIQTIYQPGSFTPLIRVETATGELA  1067

Qy    1271 IAQQRA---------------------------SKVLLTGTDSQQSVIL-----------TS  1294
                ::                          ::                     ||
Db    1068 KTQRRSLADALQQSGGEDGGSVVFPPVLVQMLDRLESEILADRVSEESRRWLASCGLTVE  1127

Qy    1295 DKQN----------------------LSQEA-------YSAYGKHKSTAND----AS  1318
                ::                     :|                |
Db    1128 QMQNQMDPVYTPARKIHLYHCDHRGLPLALISKEGTTEWCAEYDEWGNLLNEENPHQLQQ  1187

Qy    1319 ILGYNGERADPVSGVTHLGNGYRSYDPTLMRFHTPDSLSPFG-AGGINPYSYCLGDPINR  1377
                ::  |:|:           |||:               |   :  ||:
Db    1188 LIRLPGGQYDEESGLYY--NRHRYYDPLQGQRYITQD---PIGLKGGWNFYQYPL-NPVTN  1241

Qy    1378 SDPSG  1382
                ::||
Db    1242 TDPLG  1246
```

Search completed: February 9, 2005, 18:23:27
Job time : 197 secs

```
                   GenCore version 5.1.6
           Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on:     February  9, 2005, 17:59:19 ; Search time 206 Seconds
                                          (without alignments)
                                          3141.020 Million cell updates/sec

Title:          US-09-889-874A-23
Perfect score:  8879
Sequence:       1 VYIKFLKLFRRITMSDNNEF.........PRKIILGRTEKTVKPKTFRP 1673

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       2105692 seqs, 38676038l residues

Total number of hits satisfying chosen parameters:       2105692

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  A_Geneseq_16Dec04:*
            1: geneseqp19808:*
            2: geneseqp19990s:*
            3: geneseqp20008:*
            4: geneseqp20018:*
            5: geneseqp20028:*
            6: geneseqp2003aa:*
            7: geneseqp2003bs:*
            8: geneseqp20048:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                         SUMMARIES
```

Result No.	Score	% Query Match	Length	DB	ID	Description
1	8879	100.0	1673	3	AAY95707	Aay95707 Cosmid cH
2	1638	18.4	1584	3	AAY33727	Aay33727 Photorhab
3	1601	18.0	1410	8	ADQ26337	Adq26337 Chromobac
4	1583	17.8	1787	6	ABM67433	Abm67433 Photorhab
5	1569	17.7	1590	6	ABM67283	Abm67283 Photorhab
6	932.5	10.5	982	7	ADG73113	Adg73113 Pseudomon
7	932.5	10.5	982	7	ADL12168	Adl12168 Pseudomon
8	517.5	5.8	1317	8	ADS22210	Ads22210 Bacterial
9	432	4.9	2334	6	ABG31849	Abg31849 Human kin
10	386	4.3	2234	6	ABU18641	Abu18641 Protein e
11	365	4.1	1400	7	ADC01365	Adc01365 Enterohae
12	361	4.1	1510	7	ADG30698	Adg30698 Xanthomon
13	353.5	4.0	843	6	ABU48290	Abu48290 Protein e
14	350	3.9	1515	6	ABU19676	Abu19676 Protein e
15	342	3.9	1397	6	ABU15135	Abu15135 Protein e
16	341.5	3.8	1404	7	ADC00960	Adc00960 Enterohae
17	341	3.8	1394	7	ADC01426	Adc01426 Enterohae
18	340.5	3.8	1411	6	ABU15136	Abu15136 Protein e
19	339	3.8	1426	3	AAB15983	Aab15983 E. coli p
20	339	3.8	1426	6	ABU14693	Abu14693 Protein e
21	331.5	3.7	1377	4	AAU34791	Aau34791 E. coli c
22	331.5	3.7	1377	4	AAG98997	Aag98997 E. coli g
23	331.5	3.7	1377	6	ABU14807	Abu14807 Protein e
24	328	3.7	1572	6	ABU41491	Abu41491 Protein e
25	324.5	3.7	1429	6	ABM69806	Abm69806 Photorhab
26	322	3.6	1385	6	ABU40318	Abu40318 Protein e
27	317.5	3.6	1504	6	ABM66973	Abm66973 Photorhab
28	317.5	3.6	1565	6	ABU16634	Abu16634 Protein e
29	317	3.6	998	7	ABO79678	Abo79678 Pseudomon
30	314	3.5	1395	6	ABU22662	Abu22662 Protein e
31	313	3.5	1627	6	ADA35317	Ada35317 Acinetoba
32	307.5	3.5	932	6	ABU16693	Abu16693 Protein e
33	306.5	3.4	1530	6	ABU40251	Abu40251 Protein e
34	306	3.4	1586	7	ADF05044	Adf05044 Bacterial
35	298.5	3.4	1531	4	ABU21384	Abu21384 Protein e
36	298	3.4	885	6	AAU52772	Aau52772 Propionib
37	298	3.4	885	6	ABM49291	Abm49291 Propionib
38	298	3.4	920	6	ABM65127	Abm65127 Propionib
39	297	3.3	1512	6	ABU50462	Abu50462 Protein e
40	294	3.3	1439	7	ADF07275	Adf07275 Bacterial
41	292	3.3	1626	6	ABO75059	Abo75059 Pseudomon
42	291.5	3.3	1253	6	ABU40990	Abu40990 Protein e
43	291	3.3	2346	5	ADI16955	Adi16955 Murine NO
44	290.5	3.3	2802	5	ADI16953	Adi16953 Chicken N
45	289.5	3.3	1364	6	ABU47385	Abu47385 Protein e

```
                         ALIGNMENTS

RESULT 1
AAY95707
ID   AAY95707 standard; protein; 1673 AA.
XX
AC   AAY95707;
XX
DT   25-OCT-2000  (first entry)
XX
DE   Cosmid cHRIM5 encoded protein P14-2f.
XX
KW   Cosmid cHRIM5; nematocide; nematode; biological control agent;
KW   transgenic plant; helminthiasis; P14-2f.
XX
OS   Xenorhabdus bovienii.
XX
PN   WO20004285S-Al.
XX
PD   27-JUL-2000.
XX
PF   24-JAN-2000; 2000WO-GB000219.
XX
PR   22-JAN-1999;  99GB-00001499.
XX
PA   (HORT-) HORTICULTURE RES INT.
XX
PI   Morgan JAW, Jarrett P, Ellis D, Ousley MA;
XX
DR   WPI; 2000-499157/44.
DR   N-PSDB; AAA50029.
XX
PT   Novel composition used to control parasitic nematodes, especially in a
PT   plants such as maize, cotton, soya, and rice, comprises a bacterium which
PT   is a symbiont of an entomopathogenic nematode.
XX
PS   Example 6; Page 42-43; 74pp; English.
XX
CC   The present sequence is that of protein P14-2f encoded by an open reading
CC   frame identified in cosmid cHRIM5 (see AAA50029). cHRIM5 was obtained by
CC   ligating Xenorhabdus bovienii strain I73 (NCIMB 40986) Sau3A-digested DNA
CC   fragments into the BamHI site of the Stratagene cosmid vector Supercos1,
CC   packaging into Escherichia coli XL Blue 1, and screening for nematocidal
CC   activity against Caenorhabditis elegans. Analysis of the DNA indicated a
CC   number of open reading frames for which the corresponding protein
CC   sequences were determined (see AAY9568S-Y95735). Nematodes can be
CC   controlled through the use of bacteria associated symbiotically with an
CC   entomopathogenic nematode. Such bacteria include Xenorhabdus and
CC   Photorhabdus spp. such as X. bovienii strain I73. The symbiont bacteria,
CC   an engineered bacterium, or a nematocidal protein obtained from such
```

```
CC   bacteria, particularly P13-1f (see AAY95706) or P14-2f can be used to
CC   control helminthiasis in a human or domesticated animal or for the
CC   control of plant pathogen nematodes. Also claimed are vectors for
CC   expressing nematocidal proteins in host cells, and transgenic plants
XX
SQ   Sequence 1673 AA;

Query Match             100.0%;  Score 8879;  DB 3;  Length 1673;
Best Local Similarity   100.0%;  Pred. No.0;
Matches 1673;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1  VYIKFLKLFRRITMSDNNEFFTQANNFTSAVSGGVDPRTGLYNIQITLGHIVGNGNLGPT   60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  VYIKFLKLFRRITMSDNNEFFTQANNFTSAVSGGVDPRTGLYNIQITLGHIVGNGNLGPT   60

Qy     61  LPLTLSYSPLNKTDIGFGIGFNFGLSVYDRKNSLLSLSTGENYKVIETDKTVKLQQKKLD  120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  LPLTLSYSPLNKTDIGFGIGFNFGLSVYDRKNSLLSLSTGENYKVIETDKTVKLQQKKLD  120

Qy    121  NLRPEKDLKENCYRIIHKSGDIEVLTGFNNNAFDLKVPKKLLNPAGHAIYIDWNFEATQP  180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  NLRPEKDLKENCYRIIHKSGDIEVLTGFNNNAFDLKVPKKLLNPAGHAIYIDWNFEATQP  180

Qy    181  RLNRIYDDLDGHDIPLLNLEYQGLIKTILTLFPGQKEGYRTELRFLNRQLNSIHNFSLGN  240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  RLNRIYDDLDGHDIPLLNLEYQGLIKTILTLFPGQKEGYRTELRFLNRQLNSIHNFSLGN  240

Qy    241  ENPLTWSFGYTPIGKNGILGQWITSMTAPGGLKETVNYSNNNQGHHFPQSANLPVLPYVT  300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  ENPLTWSFGYTPIGKNGILGQWITSMTAPGGLKETVNYSNNNQGHHFPQSANLPVLPYVT  300

Qy    301  LMKQVPGAGOPAIQAEYSYTSHNYVGGGSNGIWNNKLDNLYGLMTEYNYGSTESSRRYDK  360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  LMKQVPGAGQPAIQAEYSYTSHNYVGGGSNGIWNNKLDNLYGLMTEYNYGSTESSRRYDK  360

Qy    361  EGHDQIVRIERTYNNYHLLTSECKQQNGYIQTTETAYYAIIGHNFDSOPSOFQLPKTKTE  420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  EGHDQIVRIERTYNNYHLLTSECKQQNGYIQTTETAYYAIIGHNFDSQPSQFQLPKTKTE  420

Qy    421  TWRSADNSYRSEITETTFDESGNPLTKVIKDKKTQKIISPSTHWEYYPPAGEVDNCPPEP  480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  TWRSADNSYRSEITETTFDESGNPLTKVIKDKKTQKIISPSTHWEYYPPAGEVDNCPPEP  480

Qy    481  YGFTRFVKKIIQTPYDSEFKDDPEKFIQYRYSLIGSQSHVTLKIEERHYSATQLLNSTLF  540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481  YGFTRFVKKIIQTPYDSEFKDDPEKFIQYRYSLIGSQSHVTLKIEERHYSATQLLNSTLF  540

Qy    541  QYNTDKSELGRLLKQTECTKGENGKTYSVVHKFYTKQDDTLQQSHSITTHDNFTIHRSQ  600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541  QYNTDKSELGRLLKQTECTKGENGKTYSVVHKFYTKQDDTLQQSHSITTHDNFTIHRSQ  600

Qy    601  VRSRYTGRLFSDTTDTKDIVTQMSYDKLGRLLTRTLNSGTPYANTLTYDYELNNLQDDNRP  660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601  VRSRYTGRLFSDTTDTKDIVTQMSYDKLGRLLTRTLNSGTPYANTLTYDYELNNLQDDNRP  660

Qy    661  PFVITTTDVNGNQLRNEFDGAGRHVSQCLKDSDGDGKFYTIHTQQYDEQGRHHTSTYSDY  720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661  PFVITTTDVNGNQLRNEFDGAGRHVSQCLKDSDGDGKFYTIHTQQYDEQGRHHTSTYSDY  720

Qy    721  LTNGRQQTDPDKVHLSMSKSYDNWGQIANTHWSYGVSEKITVDPITLTATKQLQSNSNNV  780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721  LTNGRQQTDPDKVHLSMSKSYDNWGQIANTHWSYGVSEKITVDPITLTATKQLQSNSNNV  780

Qy    781  QTGKEVTTYTPSQQPIQITLFDEAGHLQSCHTLIRDGWDRVRKETDAIGQCTIYQYDVYN  840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781  QTGKEVTTYTPSQQPIQITLFDEAGHLQSCHTLIRDGWDRVRKETDAIGQCTIYQYDVYN  840

Qy    841  RVIQITLPDGTIVNRKYAPFSTDTLITDIRVNGISLGQQTFDGLSRLTQSQDGGRVWAYT  900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841  RVIQITLPDGTIVNRKYAPFSTDTLITDIRVNGISLGQQTFDGLSRLTQSQDGGRVWAYT  900

Qy    901  YSAGNDQCPSTVITPDGQFIHYQYQPELDDAVLQVASNEITQQFSYNPVTGALLKAVAEG  960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901  YSAGNDQCPSTVITPDGQFIHYQYQPELDDAVLQVASNEITQQFSYNPVTGALLKAVAEG  960

Qy    961  QSLTPIYYPSGRLKMENINDMKKMSYLWTLRGLENGYTDLTGTIQKISRDTHGRVTQIKD  1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961  QSLTPIYYPSGRLKMENINDMKKMSYLWTLRGLENGYTDLTGTIQKISRDTHGRVTQIKD  1020

Qy   1021  SSIKTTLNYDDLNRHIGSQVTDLATGHMLTTTVBFDGLNREIGRKLCDSSGHTLDIQQSW  1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021  SSIKTTLNYDDLNRHIGSQVTDLATGHMLTTTVBFDGLNREIGRKLCDSSGHTLDIQQSW  1080

Qy   1081  LKTQQLANRIVKLNGVLQRTEQYSYDSRNRLNQYKCDGAECPTDKYGHSIVTQNFTYDIY  1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081  LKTQQLANRIVKLNGVLQRTEQYSYDSRNRLNQYKCDGAECPTDKYGHSIVTQNFTYDIY  1140

Qy   1141  GNITACHTTFADGTEDHATFKFANPTDPCQLTEVHHTHPDMPNIRLKYDKAGRVINITD  1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141  GNITACHTTFADGTEDHATFKFANPTDPCQLTEVHHTHPDMPNIRLKYDKAGRVINITD  1200

Qy   1201  NHGNTENFTYDTLGRLQNGQGSVYGYDPLNRLVSQKTDTLDCELYYRETMLVNEVRNGEM  1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201  NHGNTENFTYDTLGRLQNGQGSVYGYDPLNRLVSQKTDTLDCELYYRETMLVNEVRNGEM  1260

Qy   1261  IRLLRTGETIIAQQRASKVLLTGTDSQQSVILTSDKQNLSQEAYSAYGKHKSTANDASIL  1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261  IRLLRTGETIIAQQRASKVLLTGTDSQQSVILTSDKQNLSQEAYSAYGKHKSTANDASIL  1320

Qy   1321  GYNGERADPVSGVTHLGNGYRSYDPTLMRFHTPDSLSPFGAGGINPYSYCLGDPINRSDP  1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321  GYNGERADPVSGVTHLGNGYRSYDPTLMRFHTPDSLSPFGAGGINPYSYCLGDPINRSDP  1380

Qy   1381  SGHLSWQAWTGIGMGIAGLLLTIATGMGAIAAAGGIAAAIASTSTTALAFGALSVTSDIT  1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381  SGHLSWQAWTGIGMGIAGLLLTIATGMGAIAAAGGIAAAIASTSTTALAFGALSVTSDIT  1440

Qy   1441  SIVSGALEDASPKASSILGWVSMGMGAAGLAESAIKGGTKLATHLGAFAEDGENALLKST  1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441  SIVSGALEDASPKASSILGWVSMGMGAAGLAESAIKGGTKLATHLGAFAEDGENALLKST  1500

Qy   1501  SESSRIKWGVTRSLDREIVRNEEGQVIKDHSRGYTDNFMGKGEQAILVHGDKDGFLYHTE  1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501  SESSRIKWGVTRSLDREIVRNEEGQVIKDHSRGYTDNFMGKGEQAILVHGDKDGFLYHTE  1560

Qy   1561  GNKENGKGPYTRHTPEQLVDYLKDNNIVDLTQGGDKPVHLLSCYGKSSGAADKMAKYINR  1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561  GNKENGKGPYTRHTPEQLVDYLKDNNIVDLTQGGDKPVHLLSCYGKSSGAADKMAKYINR  1620

Qy   1621  PVIAYSNKPTISQGLARIERKDFPLKSTYHSYDPRKIILGRTEKTVKPKTFRP  1673
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621  PVIAYSNKPTISQGLARIERKDFPLKSTYHSYDPRKIILGRTEKTVKPKTFRP  1673


RESULT 2
AAY33727
ID   AAY33727 standard; protein; 1584 AA.
XX
AC   AAY33727;
XX
DT   09-NOV-1999 (first entry)
XX
DE   Photorhabdus luminescens 176 kD insecticidal toxin.
XX
KW   Symbiotic bacterium; nematode; insect; larva; toxin; insecticide.
XX
OS   Photorhabdus luminescens.
XX
PN   WO9942589-A2.
XX
PD   26-AUG-1999.
XX
PF   18-FEB-1999;  99WO-EP001015.
XX
PR   20-FEB-1998;  98US-00027080.
PR   20-JAN-1999;  99US-0116439P.
```

XX
PA   (NOVS) NOVARTIS AG.
PA   (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
XX
PI   Kramer VC, Morgan MK, Anderson AR, Hart HP, Warren GW, Dunn MM;
PI   Chen JS;
XX
DR   WPI; 1999-527479/44.
DR   N-PSDB; AAZ06826.
XX
PT   New nucleic acid from Photorhabdus luminescens encoding insecticidal
PT   toxins, used for making resistant transgenic plants.
XX
PS   Claim 25; Page 93-98; 148pp; English.
XX
CC   This sequence represents an approximately 176 kD insecticidal toxin from
CC   Photorhabdus luminescens. It is one of five insecticidal toxins (AAY33723
CC   -Y33727) encoded by open reading frames (orfs) arranged in an operon-like
CC   structure in a 9.7kb fragment of P. luminescens DNA (AAZ06826). This
CC   sequence is encoded by orf5. P. luminescens is a member of the
CC   Enterobacteriaceae family and is a symbiotic bacterium of nematodes of
CC   the genus Heterorhabditis. The nematodes colonise insect larvae, kill
CC   them, and their offspring feed on the dead larvae. However, the
CC   insecticidal agents are produced by P. luminescens rather than the
CC   nematodes. The toxins have activity against Lepidopteran insects such as
CC   Cabbage Looper (Trichoplusia ni), European Corn Borer (Ostrinia nubialis)
CC   and Fall Armyworm (Spodoptera frugiperda) and also against Coleopteran
CC   insects (e.g., Colorado Potato Beetle, Leptinotarsa decimlineata). In
CC   addition the toxins are active against strains resistant to known
CC   insecticides. The DNA sequence can be used to generate transgenic plants
CC   of various species that are resistant to economically important insect
CC   pests and also for recombinant production of the toxins for use as
CC   insecticides
XX
SQ   Sequence 1584 AA;

Query Match   18.4%;   Score 1638;   DB 2;   Length 1584;
Best Local Similarity   31.0%;   Pred. No. 4e-105;
Matches 493;   Conservative 233;   Mismatches 682;   Indels 182;   Gaps 54;

Qy   19 EFFTQANNFTSAVSGGVDPRTGLYNIQITLGHIVGN--GNLGPTLPLTLSYSPLNKTDIG 76
          : ::||||||::::||:||||| ||: |: : :||| |:|| || ||
Db   5 DIYSNAFNFGSYINTGVDPRTGQYSANINIITLRPNNVGNSEQT--LSLSFSPLTTLNNG 62

Qy   77 FGIGFNFGLSVYDRKNSLLSLSTGENYK---VIETDKTVKLQQKLDNLRFEKDLKENCY 133
          ||||| ||| :: ||:|::||:||||| :: : |||:|||| | | |||::|||:
Db   63 FGIGWRFSLTTLDIKTLTFSRANGBQFKCKCKPLPPNNNDLSFKDKKKLKDLRVYK-LDSNTF 121

Qy   134 RIIHKSGDIEVLTGFNNNAFDLKVPKKLLNPAGHAIYIDWNFEATQPRLNRIYDDLDGHD 193
          :::|:|| ::| :| | : || :|:|| ||: |: || |: :|| | ::::::
Db   122 YVYNKNGIIIEILKRIGSS--DIAKTVALEFPDGEAFDLIYNSR---------------- 162

Qy   194 IPLLNLEYQGLIKTILTL-FPGQKEGYRTELRFL-NRQLNSIHNFSLGNENPLTWSFGYT 251
          |||: ||: ||: |: |: |||:|: || || |: :||:| :|||| |: |: |
Db   163 FALSEIKYRVTGKTYLKLNYSGNN---CTSVEYPDDNNISAKIAFDYRNDYLIITVTVPYD 219

Qy   252 PIGK-------------NGILGQWITSMTAPGGLKETVNYSNNNQGHHFPQSANLPVLPY 298
          ||: ||| | | |:|:||| | :: || |:||:|| |
Db   220 ASGPIDSARFKMTYQTLKGVF-PVISTFRTPTGVELVSVYKEN--GH---KVTDTEYIPY 273

Qy   299 VTLMKQVPGAGQPAIQAEYSYTS-HNYVGGGSNGI-WNNKLDNLYGLMTEYNYGSTESRR 356
          | | :| :| |: || || ||| |::| : | | ||:|| |||||| ::
Db   274 AAALTIQPGNGQPAVSKSYEYSSVHNFLGYSSGRTSFDSSQDNLYLVTGKYTYSSIE--- 330

Qy   357 YKDKEGHDQIVRIERTYNNYHLLTSECKQNGYIQTTETAYYAIIGHNFDSQPSQFQLP- 415
          ||:|:| | |:| |:| |: |:| :|||||| | || :|||| |::||:||:|
Db   331 -RVLDGQSVVSVIERVFNKPHLMTKEAKTQDNKRITTEITYNEDLSKSFSEQPENLQQPS 389

Qy   416 KTKTETWRSADNSYRSBITETTFDESGNPLTKVIKDKKTQKIISPSTHWEYYPPAGEVDN 475
          :::| |:|| | : ||: |||||| :: || |::|:|: | | :|| |:
Db   390 RVLTRYTDIQTNTSREETVNIKSDDWGNTLL-ITETSGIQK------EYVYPVNGEGNS 442

Qy   476 CPPEPYGFTRFVKKIIQTPYDSEFKDDPEKFIQYRY----SLIGSQ------SHVTLKIEE 526
          || | : ||:|:: :| || | |:| : |:   : | |:::
Db   443 CPADPLGFSRFLKSVTQKGSPDAAQSVANKVIHYTYQKFPFTGAYVKEYVSKVSETIDN 502

Qy   527 RHYSATQLLNSTLFQYNTDKSELGRLLKQTECTKGENGKTYSVVHKFTYKQDDTLQQSH 586
          |||::| ::::|| |:: :: | : :| |: :|||:| : |||::|| || |||::
Db   503 KIARTPSYVNSP-----TSKSH-GSLAKITSVMNNQ----QTVTTFKYEYSESEMTTNA 551

Qy   587 SITTHDNFTIHRSQVRSRYTGRLFSDTDTKDIVTQMSYDKLGRLLTRTLNSGTPYANTLT 646
          |: : |||||:|| |||:|| ||| |||| |:||:| || : | ||: | ||||:| :
Db   552 TVTGFDGAHMESKNVTSIYTHRQLRKVDVNHVITDQSYDLLGRITGQIIDPGTAREIKRN 611

Qy   647 YDYELNNLQDDNRPPVITTTDVVNGNQLRNEFDGAGRHVSQCLKDSDG--------DGKF 698
          | | || : ::| :|:||::| :::| :| | : : | | ||| |||
Db   612 YVYQYPGGDENDFWP-VMIEVDSQGVRRKTHYDGMGRICSIEEQDDGAWGTSGIYQGTY 670

Qy   699 YTIHTQQYDEQGRHHTSTYSDYLTNGRQQTDP-----DKVHLSMGKSYDNWGQIANTHWS 753
          |::|:|||: || : ::| |||| || || ||: |||: | |: || |:::|
Db   671 RKVLARQYDVLGQLSKEISNDWLWN--LSANPLVRLATPLVTTKTYKYDGWGNLYSTEYS 728

Qy   754 YGVSEKITVDPITLTATK------QLQSNSNNVQTGKEVTTYTPSQQPIQITLFDEAGHL 807
          :| |:|| |::|::| | | :|:||::||: |:|||:: || |||||:| |
Db   729 DGRIELEIHDPITRTITQGVKGLGMLNIQQNNF---------EQPASIKAVYPDGTI 776

Qy   808 QSCHTLTRDGWDRVRKETDAIGQCTIYQDYNYRVIQITLPDGTIVNRKYAPFSTDTLIT 867
          || |:| ||: ::: :||:| |||||:|:|::|:::||||: :: ||:|:|:|:|:|
Db   777 YSTRTYRYDGFGRTVTETDAEGHATQIGYDVFDRIVKKTLPDGTILESAYASFSHEELIS 836

Qy   868 DIRVNGISLGQOTFDGSLRLTQSODGGRWWAYTSAGNDQCPSTVITPDGQFIHYQYQPE 927
          :| | ::| ||||:|:|| :||||: :|| ::| ::: |||||: :|:|| :
Db   837 ALNVNGTQLGALVYDGLGRVISDTVGGRKTEYLGYPQGDK-PIQSTTPSHNKQNMDYLYY 895

Qy   928 LDDAVLQVASNEITQQFSYNPVTGALLKKAVAEGQSLTPI-YYPSGRLKMEN-INDMKKMS 985
          | |::|| :|:||| ||: :||||:|: :|| : |:||: | |||:|::| |::| :|
Db   896 LGSVMSKFTTGTDQQNFRYHSKTGTLLSA-SEGVSQTNYSYFPSGVLQRESFLRDNKPIS 954

Qy   986 ---YLWTLRGLENGYTDLTGTIQKISRDTHGRVTQIKDSSIKTTLNYDDLNHIGSQVTD 1042
             |||:|||:|||||||:||||:||:||:|||:||:||:: || |:||::::| |:||
Db   955 SGEYLYTMSGLIQRHKDSFGHNHVSYDAQGRLVKTEBQDAQYATFEYDNVGRLITTTTKD 1014

Qy   1043 LATGHMLTTTVEFDGLNREIGRKLCDSSGHTLDIQQSWLKTQQLANRIVKLNGVLQRTEQ 1102
          |||| :|:|:|||||:|||:| |::|:||||||||||||:|||::|||:|||||||:||
Db   1015 TTSLSQLVTKIEYDAFDREIKRSLISDFSIQV-ITLSYTKNQISQRITSIDGVVMKNER 1073

Qy   1103 YSYDSRNRLNQYKCDGAECPTDKYGHSIVTQNFTYDIYGNITACHTTPADGTEDHATFKF 1162
          |::| |||:|:| :| ||: :||| ||: ||| | ||:|:: |:| | || :||: ||
Db   1074 YQYDNNQRLSQYCEGEQSPIDHTGRVLNQQIYHYDQWGNIKRLDNTYRDGKET-VDYHF 1132

Qy   1163 ANPTDPCQLTEVHHTHPDMPDNIRLKYDKAGRVINIT-DNHGNTENFTYDTLGRL---QN 1218
          : | | | :| :| |:|:| ||:|:: :|| | |||: ||:| | || ||
Db   1133 SQ-ADPTQLIRI----TSDKQQIELSYDANG---NLTRDEKGQT--LIYDQNNRLVQVKD 1182

Qy   1219 GQGSV--YGYDPLNRLVSQ--KTDTLDCELYRETMLVNEVRNGEMIRLLRTGETTIIAQ 1273
          | | |||: |:| :|:: :|| | ||: :|: : |:: || :|::||||
Db   1183 RLGNLVCSYQYDALNKLTAQVLANGTVNRQ-HYASGKVTNIQLGDEAITWLSSDKQRIGH 1241

Qy   1274 QRA-----SKVLLTGTDSQQSVILTSDKQNLSQEAYSAYGKHKSTANDASIIGYNGERADP 1329
          | : |:|: | ::|| ||| :| ::|: | :| | ::||:::| |:||| ||
Db   1242 QSAKNGQSVYYQYGIDHNSTVIASQNENELMALSYTPYGFRSLI---SSLPGLNGAQVDP 1298

Qy   1330 VSGVTHLGNGYRSYDPTLMRFHTPDSLSPFGAGGINPYSYCLGDPINRSDPSGHLSWQAW 1389
          |:|||:|||||||:| ||:| ||||||| | ||| :|| ||||| || ||:|: |
Db   1299 VTGWYFLGNGYRVFNPVLMRFHSPDSWSPFGRGGINPYTYCQGDPINRIDLNGHLSAGGI 1358

Qy   1390 TGIGMGIAGLLLTIAT--GGMAIA----AAGGIAAAIASTSTTALAFGALSVTSDITTSIV 1443
          ||||:|| ||||:::| ||| |:   ::| |:|| :|||:: |:| :: | : ::||:
Db   1359 LGIVLGAIGIIVGIVSLGAGAAISAGLIAAGGALGAIASTSALAVTATVIGLAADSIGIA 1418

Qy   1444 SGALEDASPKASSILGWVSMGMGAAGGLAEASAIKGGTKLATHLGAFABDGENALLKST-SE 1502
          | :| |:| :|:| :| |: | ::| :|| : :|| :| | | :|| ||:|: |
Db   1419 SAALSEKDPKTSGILNWISAGLGVLSFGISAI----------------TFT----SSLVKSARSG 1463

Qy   1503 SSRIKWGVTRSLDREIVRNEEGQVIKDHSR 1532
          |: |:| |: | | : | | |:||| ||
Db   1463 ... (continued)

```
Db   1464 SQAVSAGVIGSVPLEF-----GEVASRSSR 1488

RESULT 3
ADQ26337
ID      ADQ26337 standard; protein; 1410 AA.
XX
AC      ADQ26337;
XX
DT      23-SEP-2004  (first entry)
XX
DE      Chromobacterium violaceum CV00361 involved in hydrocyanic acid synthesis.
XX
KW      Hydrocyanic acid; biocide; pesticide; fungicide; antibacterial;
KW      nematocide; CV00361.
XX
OS      Chromobacterium violaceum.
XX
PN      WO200405696O-A2.
XX
PD      08-JUL-2004.
XX
PF      16-DEC-2003; 2003WO-BR000207.
XX
PR      19-DEC-2002; 2002BR-00007239.
XX
PA      (CNPQ-) CNPQ CONSELHO NACIONAL DESENVOLVIMENTO.
PA      (CNPQ-) CNPQ CONSELHO NACIONAL DESENVOLVIMENTO.
XX
PI      De Vasconcelos ATR, Simpson AJG, Abreu HNS, De Almeida DF;
PI      Almeida FC, De Almeida R, Antonio RV, De Araripe JR, De Araujo MFF;
PI      Bogo RM, Bonatto SL, Brigido MDM, De Brito CFA, Brochi M, Burity HA;
PI      Camargo A, Carraro D, Carvalho CMB, Cascardo JCDM, Cavada BS;
PI      Cheuire LMDO, Da Cunha MH, Fantinatti F, Farias IP, Felipe MSS;
PI      Ferrari LP, Ferro JA, Franco GR, De Freitas NSA, Furlan LR;
PI      Gattaglia D, Gazinelli RT, Gomes JAA, Goncalves PR, Grangeiro TB;
PI      Grisard EC, Guimaraes CT, Hanna ES, Jardim SN, Laurino JP, Lima LFA;
PI      De Lyra MDCCP, Madeira HMF, Maranhao AQ, Manfio GP, Martins WS;
PI      De Medeiros SRB, Meissner RDV, Moreira MA, Do Nascimento FF;
PI      Nicolas MF, De Oliveira JG, Oliveira SC, Paixao RFC, Parra J;
PI      Pasa TBC, Pedrosa FDO, Pena SDJP, Pereira JO, Pereira M, Pinto LSRC;
PI      Pinto LDS, Porto JIR, Potrich DP, Ramalho CE, Reis AMM;
PI      Rondinelli E, Sampaio AH, Dos Santos FR, Schneider MPC, Silva DW;
PI      Silva R, Soares CMDA, De Souza EM, De Souza KRL, Souza RC;
PI      Steindel M, Teixeira SMR, Trevilato PB, Urmenyi TP, Wassen R;
PI      Azevedo V, Bartoleti LA, Batista JDS, Filho AS, Zaha A;
PI      Andrade EDRM, Gonzaga L, Dos Santos EBP, Soares RDBA, Bataus LAM;
PI      Cardoso DDFDP, Parente JA, Rigo LU, Steffens MBR;
XX
DR      WPI; 2004-500292/47.
XX
PT      New gene-coding polynucleotides of the chromosome of Chromobacterium
PT      violaceum, useful for therapeutic, diagnostic or pharmacological
PT      applications, in control processes for environmental parameters or in
PT      enzyme synthesis.
XX
PS      Claim 2; SEQ ID NO 10; 31pp; English.
XX
CC      The present sequence is that of the protein encoded by the CV00361 gene
CC      of Chromobacterium violaceum strain ATCC 12472 (NCIB 9131, NCTC 9757, JCM
CC      1249, DSM 30191, IAM 12470, LMG 1267). The invention relates to 29
CC      polynucleotides that have been identified by sequencing the genome of
CC      this strain, to the polypeptides ADQ26328-ADQ26356 encoded by these
CC      polynucleotides, and to the uses of the polynucleotides and polypeptides
CC      for therapeutic, diagnostic, medicinal, pharmacological and
CC      pharmacognostic applications, in control processes for environmental
CC      parameters, and in enzyme synthetic processes. The CV00361 gene is
CC      involved in the synthesis of hydrocyanic acid. The gene and the encoded
CC      polypeptide can be used in the generation of agricultural biocides useful
CC      against bacteria, fungi and nematodes.
XX
SQ      Sequence 1410 AA;

Query Match   18.0%;  Score 1601;  DB 8;  Length 1410;

Best Local Similarity  32.2%;  Pred. No. 1.3e-102;
Matches  461;  Conservative 223;  Mismatches 577;  Indels 170;  Gaps  52;

Qy     10 RRITMSDNNEFFTQANNFTSAVSGVDPRTGLYNIQITLGHIVGNGNLGPTLPLTLSYSP 69
             : :|| ||: || ||:|||||||| :|  :
Db     22 RVLDMADNTGIYSQAFNFSSAVNGGVDPRTGLFSNFDLGILSANQGLTPALPLRIAYFP 81

Qy     70 L--NKTDIGFGIGFNFGLSV-YDRKNSLLSLSTGBNYKVIETDKT-VKLQQKKLDNLRFE 125
             :  :||:|||||||||||  :||  :||||: : :::| |: : : : || | ::  :
Db     82 LAAGESSDGLGKGFSMGGFIKYDRLRRLLILASGERYRLLDEDSAPVFIRQLKLDVVRVH 141

Qy    126 KDLKENC--YRIIHKSGDIEVLTGFNNNAFDLKVPKKLLNPAGHATYIDWNFE-ATQPRL 182
            :::|:|   |||: |:|||: |: ||||| |: ||: |  |: | :|  :|:| |  :||
Db    142 RTVADGVRRYRIVLKNGEIHHLSA--PWGPDIYPEKIQSPLGHALYLSWDSPGAGRLRL 199

Qy    183 NRIYDDLDGHDIPLLNLEYQGL--IKTILTLFP-GQKEGYRTELRFLNRQLNSIHNFSLG 239
            :: ||  :| :|     || :|   ||| |:|| | |::| ||| | |||:| |  | :|
Db    200 KEVRDE---EKRTLFRIDYPNADGERVAITQWPDSDDEKVALELYFQNGYLHRIVNKSLS 256

Qy    240 NENPLTWSFGYTPIGK--NGILGQWITSMTAPGGLKETVNYSNNNQGHHFPQSANLP--- 294
            :: ||   :|   :|:  ||:|||   : |||| |:|| :| :|:|   | |::
Db    257 GNGDVEWTLGYETDSKVADAVGGLLLNELITAPTGLAQRVRYE------PLCMKIQGER 308

Qy    295 ---VLPYVTLMKQVPGAGOPAIQAEYSYTSHNYVG---GGSNGIWNNKLDNLYGLMTE 346
                |||:| :|:|  |   | ::|||| || ::     ||: || | ||| | :||
Db    309 SDFGLPAVVLHSLVPEAGQPAINTHYEYSPANYLGYGASPKGSQG---GADELFDIQVP 364

Qy    347 YNYGSTESRRYKDKEGHDQIVRIERTYNNVHLLTSECKQQNGYIQTTETAYYAIIGHNFD 406
            |||| | || ||| ||  :| ||||| :|| :|| |  :||:|| || ||  |:|::: :
Db    365 YTYQSTEKLLDKSLNPAKPIRTTVRKYNNPHLLVSEEVREGACVFRQETAYPAKVGQSYE 424

Qy    407 SQPSQFQLPKTKTETWRSADNSYRSEITETTFDESGNPLTKVIKDKTQKIISPSTHWEY 466
            :|::| ||| || || |::: :|: :||| |||||||:  ::: |: || |:| ||:|
Db    425 AQPATFQLPVRQTMSWEAAGRSRRESIS-FEYDDAGQLIKQTMSD-------GSITVLEY 476

Qy    467 YPPAGEVDNCPPEPYGFTRFVK-KIIQTPYDSEFKDDPEKPIQYRYSLI-----GSQSHV 520
            |||||   | ||| || ||| | || ||||| || | |: |:|||: | :    |:|: |
Db    477 YPPAGEPGHCPADAEGFGFYLKSKTVFPSVSAEYGDEMAMRTEVVFRSIRTRPGSAHAEA 536

Qy    521 TLKIERHYSA-----------------------TQLLNSTLFQY--NTDKSELGRLLKQTE 557
            |||| || ||                        ||| :|| || : :||:|| | |::|
Db    537 ILQQTVSHYAGMPGPKARAAMLGGKPSWEAYQPRQLAKESYDYLDAPAQKDHGRIKKRIA 596

Qy    558 CTKGENGKTYSVVHKFTYT-----KQDDTLQQSHSITTHDN-------FTIHRSQVRS 603
            ||:|:|||:|| :|| ||      | |: ||:|| : | |:        |  : |:|
Db    597 VVYGEDRTPYEMVQDFVFEPVRSGNREVALKQTVSVTVKEDPLEKKDGKLQKVSSTRVLS 656

Qy    604 RYTGRLFSDTDTKDIVTQMSYDKLGRLLTRTLNSG-TPYANTLTYDYELNNLQDDNRPPF 662
            : |||| :|  :||::: |||||| |||:|| ||  ||::|| || ||| :|:|::|  |
Db    657 VLTGRLLSETDVLGNTVAYGYDPLGRLKTQTAHPDLKAYRAIARWDYLWPSTK--NGTPA 714

Qy    663 VITTTDVNGNQLRNEFDGAGRHVSQCLKDSDGDGKFYTIHTQQYDEQGRHHTSYSDYLT 722
            : :| ||||: |||| :|  || :|: | ||:|       ::: :  | | ||::|| |
Db    715 MAIHTDALGNQTRTSHDGLGRMIREEACDRDGGLGWKTVRTHLYDEAGRQARTTVTDVVH 774

Qy    723 NGRQGTDPDKVHLSMSKSYDNWGQIA-NTHWSYGVSEKITVDPITLT-ATKQLQSNSNNV 780
            ||:| ||| :|  |::| ||  | :|  ||:|| |:   :|||:  |  |:::|:|:  :
Db    775 D--RGGKPVTLSKTTEREWDSWGQLSVVERELETGLASRQEIDPIAQTVATWQAGTDR--- 829

Qy    781 QTGKEVTTVTPSQQPIQ--ITLFDEAGHLQS-----------CHTLTRDGWD--RVRKETDA 827
            | |:|:| | ||::|:|   || | |  | |           | | ::||| : |:  |
Db    830 CSAKYMNFYSKGSHDLERRIVL--AYHLESRSWDAEDKPYSVASWGCAHRLRRATDE 886

Qy    828 IGQCTIYQYDNYNRVIQITLPDGTIVNRKYAPFSTDTLITDIRV--NGIS-LGQQTFDG 883
            ||  | :: ::| | ||:| |||::::: :| ||  :||: ::|     ||  |:|:| ||
Db    887 MSHATGYRYDAWGRTVEIVLPDGSAVRKQYAPFSQAALPTQISVADKGVETVAGTQKFDG 946

Qy    884 LSRLTQSDQDGGRVWAYTYSAGNDQCPSTVITPDGQFIHYQYQPELDDAVLQVAS----- 937
            | | ::|:|  | |:|  :|:|:|:| | |:||:| |:| | ::| | :| ||
Db    947 LGRLKRTESGGRPTRFEYASDAASSPRTVTGPDGRVQVTAVDDRLGEALKSVAAKAPDHQ 1006

Qy    938 ---NEITQQFSYNPVTGALLKA--VAEGQSLTPIYYPSGRLKMENINDM------KKMSYL 987
                :  |||| || ||:||| |   :| |:|:||||||| |::::||: :    |||:||
```

```
Db   1007  LGVSPIQQYSYLLPMGLLHEABVGGAGQSAWD-RWPSGRLR-EETHDIRSGGKKAHYR  1064
                 |||   | |      :  |:  |     |:|| ||||   :::|
Qy    988  WTLRGLENGYTDLTGTIQKISRDT---H-GRVTQIKDSSIKTTLNYDDLNRHIGSQVTDL 1043
             ::|    |  |    | ||  |::|    ||  |||   ||
Db   1065  YSLTGNLEGGADIDGAAHARSYETAAAHVGKLIEIADAAVTVTLAYDGLQRLCSWTARD- 1123
Qy   1044  ATGHMLTTVEFDGLNREIGRKLCDSSGHTLDIQQSWLKTQQLANRIVKLNGVLQRTEQY  1103
             |||  || | |  ||  || | :|  | || || || ||
Db   1124  GRGHALATTLBFDSLGRETKKRTLAABSAEAETLSQEWYPNGQLHQRKRSEGGKPFCDETF 1183
Qy   1104  SYDSRNRLNQYKCDGAECPTDKYGHSIVTQNFTYDIYGNITACHTTFADGTBDHATFKFA 1163
             :||:||||   || |    ||:| :| :   |||  : |:  |  :|
Db   1184  VYDARNRLKDYAASGPGLPKDAYGNAIRGQKFEFDAFNNIRKCTTVLDGGSENVGEYLFE 1243
Qy   1164  NPTDPCQLTEVHHTHPD--MPDNIRLKYDKAGRVINITDNHGNTENFTYDTLGRL---QN 1218
             || ||||| |       |  ||||||||||||   : :| |||   : |:|||     :
Db   1244  NPADPCQLTKVTNSALDKGYPPAIELKYDQAGRLER--DEAG--RRLSYDALGRLARVEG 1299
Qy   1219  GQGSV-YGYDPLNRLVSQKTDT--LDCELYRETMLVNE--VRNG-------EMIRLLRT  1266
             | ||  ||||  ||:  :  |   |    ||  ||||         |     ||:|| |:
Db   1300  GGGSASYGYADHDRLVCQRVETSGMDHRLYYRANRLVNEWMTRSGQAPGADDDRVRLVYA 1359
Qy   1267  GETIIAQQR---ASKVLLTGTDSQQSVILTSDKQNLSQEAYSAYGKHKST          1313
             | :  |||    ||||:|||:|| |:|:|:|||  ||:| |||:|| | |:
Db   1360  AGSCAAQVNEGGDGSVAALMGTDGKGSIVSQAEGGQAKHYAYTPYG-HQSS          1409
```

RESULT 4
ABM67433

```
ID   ABM67433 standard; protein; 1787 AA.
XX
AC   ABM67433;
XX
DT   20-NOV-2003 (first entry)
XX
DE   Photorhabdus luminescens protein sequence #530.
XX
KW   Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KW   detection; food; gene expression; plant; animal; microorganism; toxin;
KW   antibiotic; biopesticide; virulence factor; disease model; plague;
KW   whooping cough.
XX
OS   Photorhabdus luminescens.
XX
PN   WO20294867-A2.
XX
PD   28-NOV-2002.
XX
PF   07-FEB-2002; 2002WO-IB003040.
XX
PR   07-FEB-2001; 2001FR-00001659.
XX
PA   (INSP ) INST PASTEUR.
PA   (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI   Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
PI   Buchrieser C;
XX
DR   WPI; 2003-148459/14.
XX
PT   Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
PT   useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX
PS   Claim 2; SEQ ID NO 530; 1205pp; French.
XX
CC   The invention relates to the isolation of genes and their encoded
CC   proteins from Photorhabdus luminescens. The isolated sequences are
CC   sources of probes and primers for detecting the genome of P. luminescens
CC   and related species; for gene analysis and for study polymorphisms; for
CC   detection/amplification of the genes. Antibodies (Ab) raised against the
CC   polypeptides encoded by the genes are used for detection/identification
CC   of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
CC   carry a gene-containing vector are used to select compounds that
```

```
CC   modulate, regulate, induce or inhibit expression of the genes in plants,
CC   animals or microorganisms other than P. luminescens and are able to alter
CC   response or sensitivity to toxins and antibiotics produced by P.
CC   luminescens. Cells transformed to express the genes are useful for
CC   recombinant production of the proteins, particularly toxins and
CC   antibacterials useful as insecticides, bactericides and fungicides. The
CC   genes, proteins, vectors containing the genes and Ab are also useful
CC   therapeutically (to treat microbial infection by bacteria or fungi that
CC   are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC   biopesticides. Other uses of the genes and the proteins are as virulence
CC   factors and for identifying targets of human diseases for which P.
CC   luminescens is a model (particularly plague and whooping cough). This
CC   sequence represents one of the isolated P. luminescens proteins
XX
SQ   Sequence 1787 AA;

Query Match             17.8%;  Score 1583;  DB 6;  Length 1787;
Best Local Similarity   28.2%;  Pred. No. 3.6e-101;
Matches   511;  Conservative  280;  Mismatches  727;  Indels  294;  Gaps  65;

Qy    22  TQANNFTSAVSGGVDPRTGLYNIQITLGHIVGNGNLGPTLPLTLSYSPLNKTDIGFGIGF   81
            :|| |||   :||||||||| | || || | ||| ||| | ||:|:|||| |||| ||
Db    30  SQATNFTGAMMGGVDPRTGLFTPQIPITQLGGNDLWGPDLAITLTYNPLTYLNTGYGLGF   89

Qy    82  NFGLSVYDRKNSLLSLSTGENYKVIETDKTV------KLQQKKLDNLRFEKDLKENCYRI  135
            :   :|||   :| :|:||| |:  |   |          | : |      || ||| |
Db    90  SDNFTRYDTQTQVLTLATGBIYHRVEKANEVVDGQAWTFHHAKPAHFKVKKE--KDAFWV  147

Qy   136  IHKSGDIEVLTGFNNNAFDLKVPKKLLNPAGHAIYIDWNFEATQP----RLNRIYDDLDG  191
            |:|:|  | ||| |  ||:   | :   |:|| | : | |        | ||| :
Db   148  LYKTGSREKLTQL-DRANPVAVVSEIYAPSGHKLCVKWNSFVNHNYNYWQLMEVCDAME-  205

Qy   192  HDIPILNLEYQGLIKTILTLFPGQKEGYRTELRFLNRQLNSIHNFSLGNENPLTWSFGYT  251
            ::  |  :| :|  |: :|  |   : |       |:|||||||| |||||| ||:
Db   206  ---TLLKADLATTEKIEFTWPGSPESYTVTLNMTNDLLQTVISAS-----ELTWHLEYB  257

Qy   252  PIGKNGILGQWITSMTAPGGLKETVNYSNNNQGHHFP-----------------------  288
            | |::||| |::  | | || |: || | | :|
Db   258  TEGAHKNI---LTKVTTPSGLIEKVVY--HETGHTLPTPKCIQYPNAWGPGIIRQDPKS  312

Qy   289  -QSANLPVLPVTLMKQVPGAGQPAIQAEYSYTSHNYVGGSNGIWN----NKLDNLYGL  343
             ::| :||:  | :|  | |||| |  |||||:|:| ||| |:||
Db   313  STTATVEHFPYVTQHHIIAGSGSPDQVIRYVFSPENFLGQKNKMKDPIPLPQQDNAYLA  372

Qy   344  MTEYNYGSTESRRYKDKEGHDQIVIERTYNNYHLLTSECK-----QQNGYIQTTETAY  397
            :: | | ||    |  :   |||  |:||:|| ||:|   |       | :|:|: |
Db   373  NSEYKYTSTEVREYNEKR-----YCIHREYNKFHLLVSETETVEVTPSRPQKLKETIKY  427

Qy   398  YAIIGHNFD--SQPSQFQLPKTKTETWRSADNS-----YRSEITETTFDESGNPLTKVIK  450
            ||  |  ||   || ||  | ||||| |::  :      | || | | | |||| |  |
Db   428  YADVGKSFDDNKQPPQFLMPNTVETIWHNPESSASTQRRKETTQWEYNAQCNLISMTLP  487

Qy   451  DKKTQKIISPSTHWEYPYPAGEV---DNCPPEPYGFTRFVKKI-IQTPYD-SEFKDDPEK  505
            |: | :     : :  | |:|      :|  ||:|| ||  :|     |:  ::
Db   488  DNTTTKT------TYYAPDGEETTDTHCPAEPNGFERFIKEIAVEAPSPLTKTITILR  540

Qy   506  FIQYRYSLIGSQS--------HVTLKIEERHYS-----ATQL----LNSTLFQYNTD  545
            |: :||:| || :        ||| :  ||| |     ||       || | |||:
Db   541  KVTYNYKSYDTYSPPKNNQASVVKSMVLPLSETHYSRRDCCADHLEKVKINTVSFYENTQ  600

Qy   546  KSELGRLLKQTECTKGENG-KTYSVVHKFTYTKQDDTLQQSHSITTHDNF-----TIHRS  599
            ::  ||:  | |   ||:|  |  |  :| |:|: :::   :   |::::       | :
Db   601  NAFLNGRVEQRNSYLFENGNQTRSYTEDYSWSNENKNGASCICKCKTGSGKCGTPVSHE  660

Qy   600  QVRSRYTGRLFSDTDTKDIVTQMSYDKLGRLLTRTLNSGTPYANTLTYDYELNNLQDDNR  659
            |  || ||||:  |  ::|| | ||:::||:|    :  ||:|:| | |:  ||:| ::|:
Db   661  QYWSRSTGRLLIFQKDAQDNNTVFQYDTIGHLISSTINADTAYEKTVKYAWSYSNKR----  716

Qy   660  PPFVITTDVNGNQLRNEFDGAGRHVSQCLKDSDGDGKFYTIHTQQYDEQGRHHTSTYSD  719
            | :   || :|| | |:  | |||  ::  | || |||: | :    :|  |  | |||
Db   717  --VTVTQTDIHENRYITEMDGLGRPLKKSYSPAGYQGKQFDMERYQYNPLGQLCQAISCD  774

Qy   720  YLTNGRQQTDPDKVHLSMSKSYDNWGQIANTHWSYGVSEKITVDPITLTATKQLQSNSNN  779
```

```
ABM67283
ID    ABM67283 standard; protein; 1590 AA.
XX
AC    ABM67283;
XX
DT    20-NOV-2003  (first entry)
XX
DE    Photorhabdus luminescens protein sequence #380.
XX
KW    Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KW    detection; food; gene expression; plant; animal; microorganism; toxin;
KW    antibiotic; biopesticide; virulence factor; disease model; plague;
KW    whooping cough.
XX
OS    Photorhabdus luminescens.
XX
PN    WO200294867-A2.
XX
PD    28-NOV-2002.
XX
PF    07-FEB-2002; 2002WO-IB003040.
XX
PR    07-FEB-2001; 2001FR-00001659.
XX
PA    (INSP ) INST PASTEUR.
PA    (CNRS ) CNRS CENT NAT RECH SCI.
XX
PI    Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
PI    Buchrieser C;
XX
DR    WPI; 2003-148459/14.
XX
PT    Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
PT    useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX
PS    Claim 2; SEQ ID NO 380; 1205pp; French.
XX
CC    The invention relates to the isolation of genes and their encoded
CC    proteins from Photorhabdus luminescens. The isolated sequences are
CC    sources of probes and primers for detecting the genome of P. luminescens
CC    and related species; to study polymorphisms; for gene analysis and for
CC    detection/amplification of the genes. Antibodies (Ab) raised against the
CC    polypeptides encoded by the genes are used for detection/identification
CC    of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
CC    carry a gene-containing vector are used to select compounds that
CC    modulate, regulate, induce or inhibit expression of the genes in plants,
CC    animals or microorganisms other than P. luminescens and are able to alter
CC    response or sensitivity to toxins and antibiotics produced by P.
CC    luminescens. Cells transformed to express the genes are useful for
CC    recombinant production of the proteins, particularly toxins and
CC    antibacterials useful as insecticides, bactericides and fungicides. The
CC    genes, proteins, vectors containing the genes and Ab are also useful
CC    therapeutically (to treat microbial infection by bacteria or fungi that
CC    are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC    biopesticides. Other uses of the genes and the proteins are as virulence
CC    factors and for identifying targets of human diseases for which P.
CC    luminescens is a model (particularly plague and whooping cough). This
CC    sequence represents one of the isolated P. luminescens proteins
XX
SQ    Sequence 1590 AA;

Query Match             17.7%;  Score 1569;  DB 6;  Length 1590;
Best Local Similarity   30.8%;  Pred. No. 2.9e-100;
Matches  477;  Conservative 236;  Mismatches 636;  Indels 200;  Gaps 57;
```

```
Qy       63  LTLSYSPLNKTDIGFGIGFNFGLSVYDRKNSLLSLSTGENYK---VIETDKTVKLQQKL  119
                 | |||:| |   |||| |||    :| |:  ||  ||:| |     :|    |||| |:
Db       13  LNLSFSPLTTLNNGFGIGWRFSLTMLDVKTLTPSRSNGEQFKCKPLPPNNNDISFKDKKL   72

Qy      120  DNLRFEKDLKENCYRIIHKSGDIEVLTGFNNNAPDLKVPKKLLNPAGHAIYIDWNFEATQ  179
                 ::     |:  |:.:|:.|||  ::. |:  | |:|   ||  ||:     |:
Db       73  KDLRVYK-LDSNTFYVVNKNGIIETLKRIGSS--DIAKTVALEFPDG------------  116
```

```
Db      775  HML-FENKSEKKCCSSMTVSLEYDDWGHNSRLRYSDGTSVKNTFDPIKMTSEHQRVSNDES  833
                 ::            ||:|::     |||:.       ||   ||     |||  ||  :
Qy      775  VQTGKEVTVTPSQQPIQITLFDEAGHLQSCHTLTRDGWDRVRKETDAIGQCTIYQYDNY   839

Db      834  QSSGIIRTTYNQFGQITATERLTTSRIQQGCWHYLRDELGRL-VSINANGNTTLLAYDAF  892
                 | ||||  |  |  |  ||  || :: |   | |||:||| |  :|  |||  | |||
Qy      840  NRVIQITLPDGTIVNRKYA-PFSTDTLITDIRVN--GISLGQQTFDGLSRLTQSQDGGRV  896

Qy      893  DRVIKQTFADGTTISMAYENGVSVRMSATPLGINQTPVILGTQILDGLGRVIDMESGGRK  952
                 :|||   |:  ||   |         |    :       |:.  |   |::|
Db      897  WAYTYSAGNDQCPSTVITPDGQ------FIHYQPELDDAVLQVAS--NEITQQFSYN  947

Db      953  IKLDYE-GASPVPDTVTYVKGPTSGDKDIVLHYEYEPKLNNAVTKITAGQDDVDQTFKYD 1011
                 |:|  .         |:|| |    | |     |  : |  ||    |    | | |  :
Qy      948  PVTGALLKAVAE---------GQSLTPIYPSGRLKME-----NINDMKKMSYLWTLR-G   992

Qy      993  LENGYTDLIGT--------TQKISRDTHGRVTQIKDSSIKTTLNYDDLNRHIGSQVTDL-- 1043
                 |  | |:||:        ||| ||  |  |: || :.| |:|||   |||:|||| :|
Db     1012  PKTGLLTEAIEEYKNKNDEMISSKLIFDYTLSGQLASEKLVSYNKNRAYMHFKTCYSPS  1071

Db     1072  FANRPTAIKSTINGPRYIINKYSYDNQGRLETVRNNDIEINLKYDKLSRHVCQSTEYLFD  1131
                 | | |  |  :|       ||||:  |:     |  |:  :|| |:   | |:| |
Qy     1044  ATGHMLTTVEFDGLNREIGRK------LCDSSGHT-LDIQQSWLKTQQLANRIVKLNGV  1096

Qy     1097  LQ-RTEQYSYDSRNRLNQYKCDGAECPTDKYGHSIVTQ--NFTYDIYGNITACHTTPADG 1153
                      |  ||  |.:     ||   .|| |     |  :::| |.|:  | :|  |
Db     1132  TINNKITTSLTLDDFGRETKRTIRRHNGLKNNKGITDIHISQTFNEQDKITNKEL-LHGT 1190

Db     1191  KQISKEVTYHKRGGLETYTL------MELVGEREITQCSDYKYDHLGNITQ-HSITTEG 1243
                 |::|||  ||:  ||  |        || |||: |  |   |  |  |:|   |:: |
Qy     1154  TEDHATFKFANPT--DPCQLTEVHH-THPDMPDNIRLKYDKAGRVINITD-NHGNTENFT 1209

Qy     1210  YDTLGRLQNCQGSV-------YGYDPLNRLV---SQKTDT-LDCELYYRETMLVN------ 1253
                 |  :||                 ||| |:|      | ||  | |  | |:|:
Db     1244  KTITSTYTYGNGNIQDPCQLIDVSTCTGNNSSSSLRFTYNQRGALVCENDENNTKIRTWT 1303

Db     1304  YDSLGRIDTTVDALHKVETRYLFDATNRLIIKRSEKNGTPYHHDLSYCSNSLVHDNYFYG 1363
                    |||      :     |:   |  |  |:  | :||.                :
Qy     1254  EVRNGEMIR--------LLRTGETIIAQ--QRASKVLLTGTDSQQSVILTSDKQNLSQE 1302

Qy     1303  AYSAYGKHKSTANDASILG----------YNGERADPVSGVTHLGNGYRSYDPTLM 1348
                 | |         |:        ||     |:.||   | || |  ||    || |
Db     1364  EKRDNAADRKYNKVGGICLGFSQTPCHQTPTSVASRYTETATDGKGSVIATFQGEDVQHI 1423

Db     1424  AYSPWG---VTTEQAMVTAGQQPHNTAIEEPRFNGEQWDTASASYLLGNGYRAYRPDLM 1480
                 | | .      |  ||   | |         |.| |||    |  |||| |  :|| |:
Qy     1349  RFHTPDSLSPFGAGGINPYSYCLGDPINRSDPSCHLSWQAWTGIGMGIAGLLLTIATGM 1408

Qy     1409  AIAAAGGIAAAIASTSTTALAFGALSVTSDITSIVSGALEDASPKASSILGWVSMGMGAA 1468
                 | | |  | | |:   | |  | || || | || | |:  :|   |:||| | |: :|
Db     1481  RFTAPDSWSPFGAGGINAYAYCGGSGDPVNLNDPSCGHISGWGWANIITGGIGLLLAPFTYGG 1540

Db     1541  SLEIGLGVNAARGLT------ALDAASGVTAIASGALENKNPETSRRLGWMSLGLG--  1590
                 |:| ||                 |  |  |  | |||:|  ||   || |   |
Qy     1469  GLAESAIKGGTKLA----------THLGAF----AEDGENALLKSTSB---  1502

Qy     1503  ------SSRIKWG----VTRSLDREIVRNEEGQVIKDHSRGYTDNFMGKGEQAI------L 1547
                             |     |      |      |:| | ||  :|| || |:|  |      |
Db     1591  -LPSMVIGGYSLAQWVNVRLTNSFRTPYHYPTSLGEVNLSRKSSDWVNARRSLNSGENW 1648

Db     1649  HSEVGLNGRTIWGSDTKIRGLD---IKYPLEQISRRPSNG--DIVLLSGSHGVQNGDNWL 1703
                 |:| ||  ||||| || ||           ||        :||:|
Qy     1548  VHGDKDGPLYH--------TEGNKHNGKGPYTRHTPEQLVDYLKDNNIVDLTQGGDKPV  1598

Qy     1599  HLLSCYGKSSGA 1610
                 ::|| ||  ||
Db     1704  INGSRRGSLLHFPFFKSDMTVYGGSWKGRTVHVRN-----LATMSEIDFGTLLNNGNSHI 1758

Db     1759  ILGYCYGRNDQA 1770
                 ||| ||  |:.|
```

RESULT 5

```
Qy   1131  QT--LIYDQNNRLVQVKDSKGNLVCQYQYDALNKLTAQVLANGTVNRQ-YYASGNVANVQ  1187
Db   1256  RNGEMIRLL----------RTGETIIAQQRASKVLLTGTDSQQSVILTSDKQNLSQEA    1303
Qy   1188  LGDETITWLSSDKQRLGHQSTKNGESVYYQY--------GTDHNSTVIASQNENELMALS  1239
Db   1304  YSAYGKHKSTANDASILGYNGERADPVSGVTHLGNGYRSYDPTLMRFHTPDSLSPFGAGG  1363
Qy   1240  YTPYGFRSLI---SSLPGLNGAQVDPVTGWYFLGNGYRVFNPVLMRFHSPDSWSPFGRGG  1296
Db   1364  INPYSYCLGDPINRSDPSGHLSWQAWTGIGMGIAGLLTIAT--GGMAIA-----AAGGIA  1417
Qy   1297  VNPYTYCQGDPINRIDLNGHLSAGGILGVLSAGGIIGVLGAIGIIVGVSLGAGAAISAGLIAAGGAL  1356
Db   1418  AAIASTSTTALAFGALSVTSDITSIVSGALEDASPKASSILGWVSMGMGAAGLAESAI--  1475
Qy   1357  GAIASTSAFAVTAVTVIGLAADSIGIASAALSEKDPKTAGILNWISTGLGVLSFGISAITF  1416
Db   1476  -------KGGTKLATH--LGAF-ABDGENALLKSTSESSRIKWGVTRS               1513
Qy   1417  TSSLIKSARGSGQVASTSVIGSVPIEFGEIA-----SRSSR-RWDIALS             1459
Db
```

```
RESULT 6
ADG73113
ID   ADG73113 standard; protein; 982 AA.
XX
AC   ADG73113;
XX
DT   11-MAR-2004  (first entry)
XX
DE   Pseudomonas syringae pv. tomato DC3000 Avr/Hop protein #53.
XX
KW   Avr; Hop; transgenic plant; disease resistance; cancer; bacteria;
KW   metabolic pathway; eukaryotic cell death; programmed cell death;
KW   cytostatic.
XX
OS   Pseudomonas syringae; pv. tomato str. DC3000.
XX
PN   US2003204868-A1.
XX
PD   30-OCT-2003.
XX
PF   12-FEB-2003; 2003US-00365742.
XX
PR   12-FEB-2002; 2002US-0356408P.
PR   10-MAY-2002; 2002US-0380185P.
XX
PA   (COLL/) COLLMER A.
PA   (ALFA/) ALFANO J R.
PA   (CART/) CARTINHOUR S W.
PA   (SCHN/) SCHNEIDER D J.
PA   (TANG/) TANG X.
XX
PI   Collmer A, Alfano JR, Cartinhour SW, Schneider DJ, Tang X;
XX
DR   WPI; 2003-875735/81.
DR   N-PSDB; ADG73112.
XX
PT   New nucleic acid, useful in imparting disease resistance to a plant or in
PT   preparing a composition for treating cancer.
XX
PS   Claim 15; SEQ ID NO 106; 209pp; English.
XX
CC   The present invention relates to the isolation of Pseudomonas syringae
CC   pv. tomato DC3000 Avr/Hop proteins, and the polynucleotide sequences
CC   encoding them. Also disclosed are expression vectors, host cells, and
CC   transgenic plants comprising polynucleotide sequences of the invention.
CC   The polynucleotide and polypeptide sequences are useful in imparting
CC   disease resistance to a plant or in preparing a composition for treating
CC   cancer. The sequences may also be used to make a plant hypersusceptible
CC   to colonisation by nonpathogenic bacteria, modify a metabolic pathway in
```

```
Qy   180   PRLNRIYDDLDGHDIPLLNLEYQGLIKTILTL-FPGQKEGYRTELRFL-NRQLNSIHNFS   237
Db   117   ----EVFDLIYNSRFALSEIKYRMTGKTYLKLNYSGNN---CTSVEYPDDNNISAKIAFD   169
Qy   238   LGNENPLTWSFGYTPIGK-------NGIILGQWITSMTAPGGLKETVNYSNNNQG        284
Db   170   YRNDYLITVTVPYDASGPIDSARFKMTYQTLKGIF-PVISAFRTPTGYVELVSYKEN--G   226
Qy   285   HHFPQSANLPVLPYVTLMKQVPGAGQPAIQAEYSYTS-HNVVGGGSNGI-WNNKLDNLYG   342
Db   227   HKVTDTESIPYAAALTIQ---PGNGQPAISKSYEYSSVHNFLGYSSGRTSFPDSSQDNLYL  283
Qy   343   LMTEYNYGSTESRRYKDKEGHDQIVRIERTYNNYHLLTSECKQQNGYIQTTETAYYAIIG   402
Db   284   VTGKYTYSSIE---RVLNGQNVISVTERVFDKFHLMTKEAKTQDNKRIITEITYNEDPS   339
Qy   403   HNFDSOPSQFQLPK------TKTETWRSADNSYRSEITETTFDESGNPLTKVIKDKKTQK   456
Db   340   KSFSEQPENLQQPSHVLTRYTDLQT------NTSREESVNIKSDDWGNTLL-ITETSGIQK  393
Qy   457   IISPSTHWEYYPPAGEVDNCPPEPYGFTRFVKKIIQTPYDSEFKDDPBKFIQYRYSLIGS   516
Db   394   ------EYVYPVNGEGNNCPADPLGFSRFLKSVTQKGSPDAAQSVANRVTSYTYQKLPT    447
Qy   517   QSHVTLKIEERHYSATQLINSTL---FQY---NTDKSELGRLLKQTECTKGENGKTYSVV   570
Db   448   FTGAVVK--EVVSKASETIDSKIVRTFNYVNSPTNKSH-GSLAKITSVMNNQ-----QTV   499
Qy   571   HKFTYTKQDDTLQQSHSITTHDNFTIHRSQVRSRYTGRLFSDTDTKDIVTQMSYDKLGRL   630
Db   500   TTFKYEYSDSEMTTNSTVTGFDGTHMESKNVTSIYTHRQLRKVDVNHVITDQSYDLSGRI   559
Qy   631   LTRTLNSGTPYANTLTYDYELNNLQDDNRPPFVITTDVNGNQLRNEFDGAGRHVSQCLK   690
Db   560   IGQIIDPGTTKEIKRSYIYQYPGDENDFWP-VMIEIDSQGIRRKTHYDGMGRICSIEEQ   618
Qy   691   DSDG--------DGKFYTIHTQQYDEQGRHHTSTYSDYLTN-----GRQQTDPDKVHLSM   737
Db   619   DDDGVWGTSGIYQGTYRKVLARQYDVLGQLVKEISNDWLWDLSANPLTRLTTP--LVTTK   676
Qy   738   SKSYDNWGQIANTHWSYGVSEKITVDPITLTATK-------QLQSNSNNVQTGKEVTTYTP  791
Db   677   TYQYDGWGNRYSTEYSDGRIELEIHDPITRTITQGVKGLGMLNIQQNNF----------   725
Qy   792   SQQPIQITLFDEAGHLQSCHTLTRDGWDRVRKETAIGQCTIYQYDNYRVIQITLPDGT   851
Db   726   -EQPASIKVVYPDGAIYSTRTYRYDGFGRTVTETDAEGYATQIEYDLFDRIVKKTLPDRT  784
Qy   852   IVNRKYAPFSTDTLITDIRVNGISLGQQTFDGLSRLTQSDQGRVWAYTYSAGNDQCPST   911
Db   785   ILESAYASFSHEELISALNVNGTQLGSLVVDGLGRVTRDFTVGGRKTEYLYGSQGDK-PIQ  843
Qy   912   VITP--DGQFIHYQYQPELDDAVLQVASNEITQQFSYNPVTGALLKAVABGQSLTPI-YY   968
Db   844   SVTPAHNKQNIDYLY--ALGSVMSKFTTETSQQNFSYYQKTGALLSA-TEGVSQSNYSYF   900
Qy   969   PSGRLKMENIN-DMKKMS---YLWTLRGLENGYTDLTGTIQKISRDTHGRVTQIKDSSIK   1024
Db   901   PSGVLQHESFSRDNKPISSGDYRYTMSGLIQRHKDSFAHDHVYSYDAEGRLVKTEQSSQY   960
Qy   1025  TTLNVDDLNRHIGSQVTDLATGHMLTTVEFDGLNREIGRKLCDSSGHTLDIQQSWLKTQ   1084
Db   961   ATFEYDNVGRLITTTKDTSLSQLATKIEYDVFDREIKRSLISDFSIQV-ITLSYTKNN   1019
Qy   1085  QLANRIVKLNGVLQRTEQYSYDSRNRLNQYKCDGAECPTDKYGHSIVTQNFTYDIYGNIT   1144
Db   1020  QISQRITSIDGVVMKNERYQYDSNQRLSQYCQCEGEQSPVDHTGRVLSQQIYHYDQWGNIK  1079
Qy   1145  ACHTTFADGTEDHATFKFANPTDPCQLTEVHHTHPDMPDNIRLKYDKAGRVINIT-DNHG   1203
Db   1080  RLDNTYRDGKET-VDYHFSQ-ADPTQLIRI----TSDKQQIELSYDANG---NLTRDEKG   1130
Qy   1204  NTENFYDTLGRL---QNGGQSV---YGYDPLNRLVSQ--KTDTLDCELYYRETMLVNEV   1255
```

```
CC   a cell, cause eukaryotic cell death, and inhibit programmed cell death.
CC   The present sequence represents a Pseudomonas syringae pv. tomato DC3000
CC   Avr/Hop protein of the invention.
XX
SQ   Sequence 982 AA;

Query Match             10.5%;    Score 932.5;  DB 7;   Length 982;
Best Local Similarity   29.7%;    Pred. No. 7.2e-56;
Matches   285;  Conservative 142;  Mismatches 357;   Indels 175;   Gaps    33;

Qy     642 ANTLTYDYELNNLQDDNRPPFVITTTDVNGNQLRNEFDGAGRHVSQCLKDSDGDGKFYTI 701
               :||||:  ||| ||||||||||||||||||||||:||||||:| |||
Db      31 SSTLGYRYD-----DWNQR--CCTTTDDN-VQTYEYSDPIGSDVHK---------GPIQKT 74

Qy     702 HTQQYDEQGRHHTSTYSDYLTNGRQOTDPDKVHLSMSKSYDNWGQIANTHWSYGVSEKIT 761
               |  |: |                   :|| :||              :|
Db      75 WKQSGDPEGR-----------------ISGRSET-------------WLNL------FGKPDRIR 103

Qy     762 VDPITLTATKQLQSNSNNVQTGKEVTTYTPSQQPIQITLFDEAGHLQSCHTLTRDGWDRV 821
               ||||   :  : ::||   :||    |:
Db     104 ----TLTAGKTGRSRTHSMSRSRNLTT---TEQEL------------SRQTFLYDGLGRC 144

Qy     822 RKETDAIGQCTIYQYDNYNRVIQITLPDGTIVNRKYAPFSTDTLITDIRV---NGIS--- 875
               :||  |  |||:: |||:|:: ||||  |||:|||||| ||:::|:|| |  |
Db     145 TEQRDALQQSTLFSYDNWSRMVSSTLADGSVINRSYAPQSSELATMLEVHQNGTTRTV 204

Qy     876 LGQQTFDGLSRLTQSQDGGRVWAYTYSAGNDQCPSTVITPDGQFIHYQYQPELDDAVLQV 935
               |||||||||||||:|||||||  ||| ::|    :|  | ||| ||:|| |||
Db     205 AGTQKFDGLERVTQTKTGDRVEQFNYDAGEMQ-PRSRTTAGLDNINFTYTRALTDQIFSS 263

Qy     936 ASNEITQQFSYNPVTGALLKAVAEGQSLTPIYYPSGRLKMENINDMKKMSYLW------T 989
               ::   :| :||:|| :|||     :   |
Db     264 TAPDETAKFPYDKTSARLIEATNPQGTRTRYRYDVHNQLTGETWDNL--LGQAWETRHQSS 321

Qy     990 LRGLENGYTDL------TGTIQKISRDTHGRVTQIKDSSIKTTLNYDDLNRHIGSQVTDLA 1044
               |||| | |     :|| |  :|:||||| ::| |::|| :||: |
Db     322 LLGRPIKRTDLKKGEAAGAETRYDYDTLGRIRFINQSNLRTTIDYDVLGQLCKVATEDLQ 381

Qy    1045 TGHMLTTTVEFDGLNREIGRKLCDSSGHTLDIQQSWLKTQQLANRIVKLNGVLQRTEQYS 1104
               ||:| | :| | |  :| ||::|| |:|| :|| | | ::
Db     382 AGTGVIIDMEYDDQGQEILRTQTASNQAALTLTQTWAVDGLLKTRDLQQAGSPLLHETFS 441

Qy    1105 YDSRNRLNQYKCDGAECPTDKYGHSIVTQNFTYDIYGNITACHTTPADGTEDHATFKFAN 1164
               | : | |:|  :|    :||:|||:|:|  |    :| |: |:|:|| |:|
Db     442 YDPRGRLTLVNYLGSSLPRDELQREMTRQIFSFPDELDNITLCQTRFTDGTSERAAFKYGS 501

Qy    1165 P-----TDPCQLTEVHHTHPDMPDNIRLKYDKAGRVINITDNHGNTENFTYDTLGRL--- 1216
                    |:| |  |    |||     :|    | |||  || | |  | |
Db     502 PGDDKHKDRCQLLSIAYTPRKTPDPTFSYDANGN--QLKDEHGN--SLHYDSQSRLLQV 557

Qy    1217 -QNGQG--SVYGYDPLNRLVSQKTDTLDCEL--YYRETMLVNEVRNGEMIRLRTGETII 1271
               :     :  : |::|| |: :  |::| |    | : ||: |
Db     558 AETCGAPISQYRYDGHNQLVATR-DGNESEILRFYEGHQLSSTVQEDRTQYLHLGEQPL 616

Qy    1272 AQQ----RASKVLLTGTDSQQSVILTSDKQNLSQEAYSAYGKHKSTANDASILGYNGERAD 1328
               |:     |||:||:|| :|| | |:| |:|  |   :|||  |:
Db     617 GQQIVDDAEQTLLLLTDANQSVMGEFQQGQLRKAVYSAYGERHSEEALLSTAGFNGEVRE 676

Qy    1329 PVSGVTHLGNGYRSYDPTLMRFHTPDSLSPFGAGGINPYSYCLGDPINRSDPSGH----- 1383
               | :  |||||||    :|: :|  || ||  ||| || |  |||
Db     677 AANGWYLLGNGYRAYNPLLMRFHSPDFLSPFAEGGVNPYTYCLGNPIALRDPTGHDASCQ 736

Qy    1384 ------LSWQ-------AWTGIGMGIA---GLLLTIATGGMAIAAAGI----------- 1416
                    |:|         |   |  |      ::|:|::|  | |||:|
Db     737 TGRLRRPDEGALPMQQGGGDIMGWVGVGIGVVF-VLGVAATIATLGTATPVTGPVTVLGI 796

Qy    1417 --------AAAIASTSTTALAFG----ALSVTSDITSIVSGALEDASPKASSILGWVS---- 1462
                        ||||::|: ||     || |:|| |  :| |:|||
Db     797 SMTASAAAVSTVSTGALIVGTALTAASTTANTVAIVNN-----DQTAGEVGWLGIAAV 851

Qy    1463 ------MGMGAAGLAESAIKGGTKLA------------THLGAFAEDGENALLKSTSESSR 1505
                    |||| | |||| :||| |           ||||| ||: ::: |: |::|
Db     852 PVGLVGFGAGAVVARAVARAVAAAAKVAAANAGTIGVRSVSRIG-LAAAGARRTISSAASSAR 909
```

```
PA   (CAOV/) CAO Y.
PA   (HINK/) HINKLE G J.
PA   (SLAT/) SLATER S C.
PA   (CHEN/) CHEN X.
PA   (GOLD/) GOLDMAN B S.
XX
PI   Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX
DR   WPI; 2004-061375/06.
XX
PT   New recombinant DNA construct comprising a promoter positioned to provide
PT   for expression of a polynucleotide encoding a polypeptide from a
PT   microbial source, useful for producing plants with improved properties.
XX
PS   Claim 1; SEQ ID NO 11243; 122pp; English.
XX
CC   The invention relates to a recombinant DNA construct comprising a
CC   promoter functional in a plant cell, where the promoter is positioned to
CC   provide for expression of a polynucleotide encoding a polypeptide from a
CC   microbial source. The invention also relates to a transformed plant
CC   comprising the recombinant DNA construct and a method of producing a
CC   transformed plant having an improved property. The plant is a crop plant
CC   such as maize or soybean. The method of producing a transformed plant
CC   having an improved property comprises transforming a plant with the
CC   recombinant DNA construct and growing the transformed plant, where the
CC   polynucleotide or polypeptide is useful for improving plant properties.
CC   The recombinant DNA construct is useful for producing plants with
CC   improved plant properties, e.g. improved cold, heat or drought tolerance,
CC   tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
CC   increased resistance to plant disease, better growth rate by modification
CC   of the cell cycle pathway with plant growth regulators, increased rate of
CC   homologous recombination, modified seed oil or protein yield and/or
CC   content, improved yield by modification of carbohydrate, nitrogen or
CC   phosphorus use and/or uptake, by modification of photosynthesis or by
CC   providing improved plant growth and development under at least one stress
CC   condition, improved lignin production or improved galactomannan
CC   production. This sequence represents a bacterial polypeptide used in the
CC   scope of the invention. Note: The sequence data for this patent did not
CC   form part of the printed specification but was obtained in electronic
CC   format from USPTO at seqdata.uspto.gov/sequence.html.
XX
SQ   Sequence 1317 AA;

Query Match             5.8%;  Score 517.5;  DB 8;  Length 1317;
Best Local Similarity   21.2%;  Pred. No. 1.7e-26;
Matches  334;  Conservative 194;  Mismatches 505;  Indels 541;  Gaps  72;

Qy   39 TGLYNIQITLGHIVGNGNLGPTLPLTL--SYSPLNKTDIGFGIG--FNFGLSVYDRKNSL 94
         ||  |  : :||  |:  |:  : |:|:||   |:  ||||| |  | :|  :||:|:|
Db   56 TGSYFYQYQDLSIPGRG-----LPLTVSRSYNSMDNRSGLFGSGWTFNYNMNLTVDNNGN 110

Qy   95 LSLSTGENYK---VIETDKTVKLQQKKLDNLRFEKDLKENCYRIIHKSGDIEVLTGFNNN 151
         :|:   ||||     |  | : : :|        ||   |      :|    |:|
Db  111 VTVLGGDGHTDTYIILNPDGTYSRPLSVFDDL-----------IKNSDETYTLTKDQT 157

Qy  152 AFDLKVPKKLLN--PAGHAIYIDWNFEATQPRLNRIYDDLDGHDIPLLNLSYQGLIKTI 208
         |  ||  |:||    | ||  :||        | : ||| | |:|  |:|:||| |:
Db  158 KYNFSSEGKLVNIVDKNGNQI---NFTYTGEQLTKV-TDASGREL-ILAYDHNG------ 206

Qy  209 LTLFPGQKEGYRTELRFLNRQLNSIHNFSLGNENPLTWSFGYTPIG-----KNGILGQ-- 261
         |:|  |:|      |  | :||||  ||  ||  |||: |  ||||       |  ||
Db  207 -----------------------HIISITDPMGRVWSYSYDDQGNLIQCKNPIGGKLS 241

Qy  262 -------WITSMTAPGG---LKETVNYSNNNQGHHFPQSANLPVLPYVTLMKQVPGAGQP 311
               | :||| |||       | :  : || : |  |    : |    :|  |:|  |
Db  242 YTYDENHWMTSITDPRGNHPMKNTY----DEKGHVISQS------------------- 276

Qy  312 AIQAEYSYTSHNYVGGGSNGIWNNKLDNLYGLMTEYNYGSTESRRYDKEGHDQIVRIER 371
         ||            ||::|| :|   : :  | : |  ::   :|
Db  277 -----------------NSLNATY---TFNYDS-ENRK--------------------- 293

Qy  372 TYNNYHLLTSECKQQNGYIQTETAYYAIIGHNFDSQPSQFQLPKKTKTETWRSADNSYRS 431
         ||||                :  ||     :| |  :       ||
```

```
Qy  936 ASNEITQQFSYNPVTGALLKAVAEGQSLTPIYYPSGRLKMENINDMKKMSYLW------T 989
         : :| |  |   |||| ||:  |      : : |||||:|:|  :::
Db  264 TAPDETAKFDYKTSARLIEATNPQGTRTYRYDVHNQLTGSTWDNL--LGGAWETRHQSS 321

Qy  990 LRGLENGYTDL------TGTIQKISRDTHGRVTQIKDSSIKTTLNYDDLNRHIGSQVTDLA 1044
         | | ||||         |   |: |   || :: ||||:|:|||||| |   |  | ||
Db  322 LLGRPIKRTDLKKGEAAGAETRYDYDTLGRIRRFINQSNLRTTIDYDVLGQLCKVATEDLQ 381

Qy 1045 TGHMLTTTVEPDGLNRBIGRKLCDSSGHTLDIQQSWLKTQQLANRIVKLNGVLQRTEQYS 1104
         :|    :  : ||:|||||:| |:||| | |: :  : ::  ||:::  :| | ::  |:
Db  382 AGTGVIIDMEYDDQGQEILRTQTASNQAALTLTQTWAVDGLLKTRDLQQAGSPLLHETFS 441

Qy 1105 YDSRNRLNQYKCDGABCPTDKYGHSIVTQNFTYDIYGNITACHTTFADGTEDHATFKFAN 1164
         ||  :||  | ||   | : :: | :: :| |   |:   |  | | ||:| :| ||:
Db  442 YDPRGRLTLVNYLGSSLPRDELQREMTRQIPSFDELDNITLCQTRFTDGTSERAAFKYGS 501

Qy 1165 P------TDPCQLTEVHHTHPDMPDNIRLKYDKAGRVINITDNHGNTENFTYDTLGRL--- 1216
         |       | |:|::   :  |: :   | || |  :: |::||| |::  |||:  |
Db  502 PGDDKHKDRCQLLSIAYTPPRKTPDPTFSYDANGN--QLKDEHGN--SLHYDSQSRLLQV 557

Qy 1217 -QNGQG--SVVGYDPLNRLVSQKTDFLDCEL--YRRETMLVNEVRNGEMIRLLRTGETII 1271
           | |    : :|:|| :  |:     :    |  | :   ::  :  :  |  | :
Db  558 AETGGAPISQYRYDGHNQLVATR-DCNESEILRFYEGHQLSSTVQEDQRTQYLHIGEBQPL 616

Qy 1272 AQQ------RASKVLLTGDSQQSVILTSDKQNLSQEAYSAYGKHKSTANDASILGYNGERAD 1328
         |||        :  |||| ::|  :|:: :  |  ||:| |:| :    |   |  |   :
Db  617 GQQIVDDAEQTLLLLTDANQSVMGEFQQGQLRKAVTSAYGERHSEEALLSTAGFNGEVRE 676

Qy 1329 PVSGVTHLGNGYRSYDPTLMRFHTPDSLSPFGAGGINPYSYCLGPDPINRSDPSGH---- 1383
         | .| :|||||| :|:|  |||| ||.||||  |:|||:|| |   | :|||:|
Db  677 AANGWYLLGNGYRAYNPLLMRFHSPDFLSPFAEBGGVNPYTYCLGNPIALRDPTGHDASGQ 736

Qy 1384 -------LSWQ----AWTGIGMGIA----GLLLTIATGMAIAAAGGI--- 1416
                |:|      | |||| ||     |::  :| |   | :
Db  737 TGRLRRPDEGALPMQQGGGDIMGWVGVGIGVVFTVLGVAATIATLGTATPVTGPVTVLGI 796

Qy 1417 -----AAAIASTSTTALAPG---ALSVTSDITSIVSGALEDASPKASSILGWVS----- 1462
              |||:|:||::| | |     || |: |:||||:  : : ||   |||
Db  797 SMTASAAAVSTVSTGALIVGTALTAASTTANTVAIVNN----DQTAGEVGGWLGIAAV 851

Qy 1463 ----MGMGAAGLAESAIKGGTKLA-----------------THLGAPAEDGENALLKSTSESSR 1505
              :| | || |:|:  | :|:|                  ||| |  :  | :|  : ||| |
Db  852 PVGLVGFGAGAVVARAVARAVAAAAKVAAANAGTIGVRSVSRIG-LAAAGARRTISSAASSAR 909
```

RESULT 8
ADS22210

```
ID   ADS22210 standard; protein; 1317 AA.
XX
AC   ADS22210;
XX
DT   02-DEC-2004  (first entry)
XX
DE   Bacterial polypeptide #11243.
XX
KW   Recombinant DNA construct; transformed plant; improved plant property;
KW   cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
KW   pathogen tolerance; pest tolerance; plant disease resistance;
KW   cell cycle pathway modification; seed oil yield; protein yield; carbohydrate;
KW   homologous recombination; seed oil yield; protein yield; carbohydrate;
KW   nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
KW   bacterial polypeptide.
XX
OS   Bacteria.
XX
PN   US2003233675-A1.
XX
PD   18-DEC-2003.
XX
PF   20-FEB-2003; 2003US-00369493.
XX
PR   21-FEB-2002; 2002US-0360039P.
XX
```

```
Db    294  -------------------------------------------TTETD------------------------------- 313

Qy    432  EITET--------TFPDESGNPLTKVIDKDKKTQKIISPSTHWEYYPPAGEVDNCPPEPYG 482
                :||:|        :|  :|  ::|||::|
Db    314  ELNETNQLGVTISYAYDENGNRISVTNENSKTTKLA--------YDANGNIIK-TTNPLG 364

Qy    483  FTRFVKKIIQTPYDSEFKDDPEKFIQYRYSLIGSQSHVTLKIEERHYSATQLLNS------- 537
                :::|                  ||
Db    365  YSK------SMTYDS--------KNNLISQITDELRHKTSFEYDDNSNLIKSIDALG 406

Qy    538  --TLFQYNTDKSELGRLLKQTECTK-----------GENGK-----TYSVVHKFT 574
             ||:|:           |::  ||||
Db    407  HETVFSYD---KYGQVIGETDSNKKTATFSYNNNGDQITITDANGKTSAFTYDTVGRVT 462

Qy    575  ----------VTKQDDTLQQSHSITTHDNFTIHRSQVRSRYTGRL--------FSDTDTKDI 618
                      ||
Db    463  TKTDAKGNRYTFQYDALDNLLSITD------PMGQTTSTNTMLLEIKLVLQMLKVDXQSI 516

Qy    619  VTQMSYDKLGRLLTRTLNSGT------PYA---------NTLTYDYELNNLQDDNRPPFVIT 665
                :::|                         :  :|||:
Db    517  LTTLXISXLXKRMQWVEXLGTNMMPLAIWFPRQTQKGHKTSYDYDPLNRQ--------VS 568

Qy    666  TTDVNGNQLRNEFDGAGRHVS--------------QCLKDSDGDGKFYTIHTQQYDE 708
                ||                          ::
Db    569  VTNALGKTTRNKYDAIGNKISITNAYGKSTRYSYNSLNQLVKVTNAMGK---VVRYNYDA 625

Qy    709  QGRHHTSTYSD--------YLTNGRQQTDPDKVHLSMSKSYDNWGQIANTHWSYGVSEKIT 761
                ::|:       ||
Db    626  VGNLISTTDENGHKINYGYDSLNRQVSVTDALRKTTRNKYDAVGNKISITNAYGKSTRYS 685

Qy    762  VDPITLTATKQLQSNSNNVQTGKEVTTYTPSQQPIQITLFDEAGHLQSCHTLTRDGWDRV 821
                 |:           :  :    |||  |:||:
Db    686  Y-------NSLN-----QLVKVTDAMGGVVRYNYDAVGNLIS------------------ 715

Qy    822  RKETDAIGQCTIYQDYNYNRVIQITLPDGTIVNRKYAPFSTDTLITDIRVNGISLGQQTF 881
                |                ::|  ||
Db    716  --TTDANGRKTNYGYDSLNRQVSITNALGKTRNKYDAVGNKISSTDANWR---LTKYSY 770

Qy    882  DGLSRLTQSODG-GRVWAYTYSA-GNDQCPSTVITPD--GQFIHQY-----QPELDDAV 932
                ||                     ||         |:||
Db    771  DSLNRLVKVTDAMGGVVRYTYDAVGN--------LISTTDAKGHKITDYGYDSLDRQVSITDPL 825

Qy    933  LQVASNEITQQFSYNPVTGALLKAVAEGQSLTPIYYPSGRLKMENINDMKKMSYLWTLRG 992
                :  ||        |::|                   |:||:|
Db    826  GRTARNK------YDAVGNKISSTDEDGKTTSVGYDVLNRLTKVSYPDDQKVSY------- 873

Qy    993  LENGYTDLTGTIQKISRDTHGRVTQIKDSSIKTLNYDDLNRHI------GSQV--TDLA 1044
                                                       ||
Db    874  --------------------NYDAVGNRLTMKDSHGTTAVKYDKLNRLLSVLNPDGGQKVSYTYNK 918

Qy    1045 TGHML--------TTTVEFDGLNREIGRKLCDSSGHTLD--IQQSWLKTQQLANRIVKL 1093
                 |:|                 ||::|
Db    919  VGNRVKMTYPDGKTTSYSYDAVNRLIG--VIDSDGHITSYSYAKGNLKTMNTPNGV--- 973

Qy    1094 NGVLQRTEQYSYDSRNRLNQYKCDGAEPCTDKYGHSIVTQNFTYDIYGNITACHTTPADG 1153
                |:                                      ||
Db    974  -----KTE-YSYDKANRL--------VELINKNTQVVSSYKYTLDAAGNRLKVDEQLAEG 1020

Qy    1154 TED----------HATFKFANPTDPCQLTEVHHTHPDMPDN--IRLKYDKAGRVINIT- 1199
                                    :|          ::
Db    1021 VESGDSELKESQLLTTYGY---DKLYRLJTKV-----DYPSNKTVSYKYDSMGNRISMTT 1072

Qy    1200 --DNHGNTENFTYDTLGRLJQNGQGSVYGYDPLNRLVSQKTDTLDCELY-YREFMLVNEVR 1256
                :::                |:|:|
Db    1073 NVDGIGSTISYKYDAADQLLQSGNISYSYDKKGNLIIKKRVNSTQFMSYSYDE--ANRLK 1129

Qy    1257 N-----------GEMIRLLRT--GETIIAQQRASKVLLTGTDSQQSVILT 1293
                                ::|:|
Db    1130 NVSEFVSNTNTPKSSYNFEYDGDGNRIIKTTINGE----NAQSTKYVLDINSALPQVLTE 1185

Qy    1294 SDKQNL-------SQEA------YSAYGKHKSTANDASIL------------ 1320
             |:||:|       |:||
Db    1186 SDTKNTTCYTYGTDLISMTNSENAEYYYHYDGLGSVRSLSDSKGIIKNTYLYDAFGQVQK 1245
```

```
Qy    1321 --------GYNGERADPVSGVTHLGNGYRSYDPTLMRFHTPDSLS--PFGAGGINPYSY 1369
                   :|:  | ::|:       |:|  |||:
Db    1246 EIGTVDNEFPTGEOMDDETGLIYLRARY--YDPSVGRFITKDVIKGRRVTTQSINRYVV 1303

Qy    1370 CLGDPINRSDPSGH 1383
                :|:  |
Db    1304 TTNNPVNLVDLTGY 1317
```

RESULT 9

ABG31849

ID   ABG31849 standard; protein; 2334 AA.
XX
AC   ABG31849;
XX
DT   05-NOV-2002 (first entry)
XX
DE   Human kinase, MEK1.
XX
KW   Extracellular signal regulated kinase; hyperalgesia; surgery;
KW   opioid withdrawal; pain sensitisation; analgesic; chronic pain; ERK;
KW   MEK1; human; enzyme.
XX
OS   Homo sapiens.
XX
PN   WO200258687-A2.
XX
PD   01-AUG-2002.
XX
PF   25-JAN-2002; 2002WO-US002128.
XX
PR   25-JAN-2001; 2001US-0264336P.
XX
PA   (TEXA) UNIV TEXAS SYSTEM.
XX
PI   Gutstein HB;
XX
DR   WPI; 2002-608414/65.
DR   N-PSDB; ABK90804.
XX
PT   Reducing or reversing tolerance, physical dependence, hyperalgesia,
PT   withdrawal symptoms, or pain sensitization in patients on analgesics for
PT   chronic pain, comprises inhibition of the extracellular signal-regulated
PT   kinase (ERK).
XX
PS   Disclosure; Page 156-161; 163pp; English.
XX
CC   The invention relates to a method of reducing or reversing tolerance,
CC   reducing the risk of physical dependence or hyperalgesia, reducing the
CC   symptoms of opioid withdrawal or inhibiting pain sensitisation in a
CC   patient taking analgesics. The method comprises administering an
CC   analgesic and an extracellular signal-regulated kinase (ERK) inhibitor
CC   comprised in a formulation to reduce or reverse tolerance, risk of
CC   physical dependence, hyperalgesia, symptoms of opioid withdrawal, or
CC   inhibiting pain sensitisation in patients taking analgesics for chronic
CC   pain or those undergoing surgery. The present sequence represents the
CC   amino acid sequence of human MEK1 (not defined).
XX
SQ   Sequence 2334 AA;

```
Query Match             4.9%;  Score 432;  DB 5;  Length 2334;
Best Local Similarity  20.4%;  Pred. No. 4.5e-20;
Matches  358;  Conservative  226;  Mismatches  613;  Indels  560;  Gaps  80;

Qy     44  IQITLGHIVG-NGNL-----------------GPTLPLTLSYSPLNKTDIGEIGFNFGLSVVDR 90
               |                            ::||      :||
Db    896  IDIPSGQLNGATGNVIVNEEDLSIDGRGPGLGLSRTNSLDSSDHLFGQGW----YADA 950

Qy     91  KNSLLSLSTGENYKVIETDKTVKLQQKKLDN-----LRFEKDLKENCYRIIHKSGDIEVL 145
                ::||:                            :::|   |:
Db    951  ETSVISTDGGAMY--IDEDATHRFTKADGTYQPPTGVYLELTETADQFILKTKD---- 1004
```

```
Db   1899 ------------------------------------ASGKV----QYEYGKLNQLVK-- 1927
          :::|:                                       ||| |
Qy   1117 DGAECPTDKYGHSIVTQNFTYDIYGNITACHTTFADGTED--HATFKFANPTDPCQLTEV 1174
Db   1928 -----EYTYDGFGNRKTV-TTIKDGSSKTVNASFNIMN-----QLTKV 1973
          :|||   ||:|||
Qy   1175 H------HTHP-DMPDNI------------------RLKYDKAGRVINITDNHG 1203
Db   1974 NDESISYDKNGNRTSDGKFTYTWDAEDNLTAVTKKGEDKPFATYKYDEKGNRIQKTVN-G 2032
          :::||
Qy   1204 NTENFTYDTLGRLQNGQGSVVGYDPLNRLVSQKTDTLDCELY-YRETMLVNEVRNGEMIR 1262
Db   2033 KVTNYFYDG-------------------DSLNVLYETDADNNVTKSYTYGD---------SGQLLS 2070
          ::|||                      :||
Qy   1263 LLRTGETIIAQQRASKVLLTGTDSQOSVILTSDKQNLSQEAYSAYGK-HKSTANDA----S 1318
Db   2071 YTENGKKYFYHYNAHGDIIAISDSTGKTV-------AKYQYDAWGNPTKTEASDEVKDN 2122
          ::                              :
Qy   1319 ILGYNGERADPVSGVTHLGNGYRSYDPTLMRHT--PDSLSPFGAGGINPYSYCLGDPIN 1376
Db   2123 RYRYAGYQYDEETGLYYLMARY--YEPRNGVFLSLDPDPGSDGDSLDQNGYAYGNNNPVM 2180
          ::|:                      :|||
Qy   1377 RSDPSGHLSWQAWTGIGMGIAGLLLTIATGGMAIAAAG---GIAAAIASTSTTALAFGA 1432
Db   2181 NVDPDGH--W-VWLVVNAGFA------AYDGYKAYKSGKGWKGAAWAAAASNFGPGKIFKG 2231
          ::||
Qy   1433 LSVTSDITSIVSGALEDASPKASSILGWVSMGM----GAAGLAEBASAIKGGTKLATHLGAF 1488
Db   2232 ASRAYKFTK-----KAVKITGHTRHGLNQSIGRNG------GRGVNLRAKLNA- 2273
          ::|                     :|
Qy   1489 AEDGENALLKSTSESSRIKWGTVTRSLDRE--IVRNEEGQVIKDHSRGYTDNFMGKGBQAI 1546
Db   2274 ------------VRSPKKVIKQPNGATKYVGKKATVVLNKRGKVITAYG-----------------SS 2312
          ::                                             ||
Qy   1547 LVHGDKDGFLYHTEGNK 1563
Db   2313 RAKGSKHFVFHTHGKGNK 2329
```

RESULT 10
ABU18641
ID   ABU18641 standard; protein; 2234 AA.
XX
AC   ABU18641;
XX
DT   19-JUN-2003 (first entry)
XX
DE   Protein encoded by Prokaryotic essential gene #4168.
XX
KW   Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS   Bacillus anthracis.
XX
PN   WO200277183-A2.
XX
PD   03-OCT-2002.
XX
PF   21-MAR-2002; 2002WO-US009107.
XX
PR   21-MAR-2001; 2001US-00815242.
PR   06-SEP-2001; 2001US-00948993.
PR   25-OCT-2001; 2001US-0342923P.
PR   08-FEB-2002; 2002US-00072851.
PR   06-MAR-2002; 2002US-0362699P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Wang L,   Zamudio C,   Malone C,   Haselbeck R,   Ohlsen KL,   Zyskind JW;
PI   Wall D,   Trawick JD,   Carr GJ,   Yamamoto R,   Forsyth RA,   Xu HH;
XX
DR   WPI; 2003-029926/02.
DR   N-PSDB; ACA22511.

```
Qy    146 TGFNNNAFDLKVPKKLL------NPAGHAIYIDWN-----FEATQPRLNRIYDDLDGH-- 192
          :|:||
Db   1005 ---QTNAYFNKKGGKLQKVVDGHNNATVYTVYTNDKNQLTAITDASGRKLTFYDE-NGHVT 1060
Qy    193 ------------------DIPLLNLEY--QGLI-----------KTILTLFPG 214
          :|:
Db   1061 SITGPKNKKVTYSYENDLLLKKVTDTDGTVTSYDYDSEGRLVKQYSANSTEAKPVFTEY-- 1118
Qy    215 QKEGYRTELRFLNRQLNSIHNFS-------LGNENPLTWSFGYTPIGKNGILGQWITSMT 267
          ::|:: |::|
Db   1119 QYSGHRLEKAINAKKETYVYSYDADKKTLLMTQPNGRKVQYGYNEAGNP------IQVID 1172
Qy    268 APGGLKETVN--YSNNN-----------QGHHFPQSANLPVLPYVTLMKQVPGAG 309
          :|:
Db   1173 DAEGLKITTNTKYEGNNVVEDVDPNDVGTGKATESYQYDKDGN-----VTSVKDAYGT-- 1225
Qy    310 QPAIQAEYSTHSNYVGGGSNGIWNNKLDNLYGLMTEYNYGSTESRRYKDKEGHDQIVRI 369
          ::|
Db   1226 ----ETYEYNKNNDV-----TKMKDTEGNVTDIAYDGLDAVSETDQSGKSSSAAV 1271
Qy    370 ERTNNYHLLTSECKQONGYIQTTETAYYAIIGHNFDSQPSQFQLPKTKTETWRSADNSY 429
          :|||: ||:
Db   1272 YDKYGNQIQSSKDLSASTNILK----------DGSFEAQKSGWNLTASKD------ 1311
Qy    430 RSEITETTFDESGNPLTKVIKDKKTQKIISPS----THWEYYPPAGEVDNCPPEPYGFTR 485
          :|| :
Db   1312 RRKIS-VIADKSG-----VLSGSKALEVLSQSTSAGTDHGYSSATQTVELEPNTTYTLSG 1365
Qy    486 FVKKII---QTPYDSEFKDDPEKFIQY---RYSLIGSQSHVTLKIEERHYSATQLLNS-- 537
          ||:|:                      ||
Db   1366 KIKTDLAKSRAYFNIDLRDKDQKRIQWIHNEYSALAGKNDWT----KRQITFTTPANAGK 1421
Qy    538 ---TLFQYNTDKSELGRL-LKQTECTKGENGKTYSVVHKFTYTKQDDTLQQS-HSITTHD 592
          :||:: |: |:
Db   1422 AVVMEVDHKDKDGKGKAWFDEVQLEKGEVSSSYNPVQNSSFTSATENWNVSGASVDSEE 1481
Qy    593 NFTIHRSQVRSRYTGRLFSDTDTDKDIVTQMSYDKLGRLLTRT-----------L 635
          :|:|
Db   1482 GFNDDVSLKAARTSASQAGSVTKQTVVLGQSANDKPVYLTLTGMSKASSVKFTDEKDYSL 1541
Qy    636 NSGTPYANTLTYDYEL--NNLQDDNRPPFVI 664
          ::||
Db   1542 QANVTYADGSTGIYNAKFPSGTQEWNRAAVVIPKTKPINKVDISILFQKSATGTVWFDDI 1601
Qy    665 ------------TTTDVNGNQLRNEFDGAGRHVSQCLKDSDGDGKFYTIHTQQYDEQGRHHT 714
          ::|||:
Db   1602 RLIEGSLLLTKSTYDSNGNYVTKEEDELGYATS---TDYDETGK------KTSETDAKGEKTT 1654
Qy    715 STY--SDYLTNGRQQTDPDKVHLSMSKSYDNWG-QIANT--------HWSYGVSEKI-- 760
          ::|:: |||
Db   1655 YTTYDQADQLTNMTLSNGTSILH-----SYDKEGNEVSKTIRAGADQTYKFEYDVMGKLVK 1709
Qy    761 TVDPITLTATKQLQSNSNNVQT---GKEVTTYTPSQQPIQITLFDEAGHLQSCHTLTRD 816
          ::|
Db   1710 TTDPLGNVLASEYDANSNLTKTISPNGNEV---------SLSYD 1744
Qy    817 GWDRVRKETDAIGQCTIYQDYNYNRVIQITLPDGTIVNRKYAPFSTDTLITDIRVNGISL 876
          ::|: |:
Db   1745 GTDRVKSKSYNGTEKYIFFTYDKGN------ETSVVNKEQN--TT 1781
Qy    877 GQQTPDGLSRLTQSODGGRVWAYTYSAGNDQCPSTVITPDGQFIHYQYQPELDDAVLQVA 936
          :|:|:
Db   1782 KKRTFDNKNRLTELTDRGGSQTWTYPSDSDKLKTF------SWIH---------G 1821
Qy    937 SNEITQQFSYNPVTGALLKAVAEGQSLTPIYYPSGRLKMENINDMKKMSYLWTLRGLENG 996
          ::: |: |:
Db   1822 DQKGTNQFTYN-------------------KLDQMIEMKDSTSSYSFDYDEN- 1854
Qy    997 YTDLTGTIQKISRDTHGRVTQIKDSSIKTLNYDDLNRHIGSQVTDLATGHMLTTTVEFD 1056
          ::|||:
Db   1855 ------GNVQ--------TFITGNGGGTSFSYDERNLVSSLHIGDKNGGDILTESYEY- 1898
Qy   1057 GLNREIGRKLCDSSGHTLDIQQSWLKTQQLANRIVKLNGVLQRTEQYSYDSRNRLNQYKC 1116
```

```
XX
PT   New antisense nucleic acids, useful for identifying proteins or screening
PT   for homologous nucleic acids required for cellular proliferation to
PT   isolate candidate molecules for rational drug discovery programs.
XX
PS   Claim 25; SEQ ID NO 46565; 1766pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising any one of
CC   the 6213 antisense sequences given in the specification where expression
CC   of the nucleic acid inhibits proliferation of a cell. Also included are:
CC   (1) a vector comprising a promoter operably linked to the nucleic acid
CC   encoding a polypeptide whose expression is inhibited by the antisense
CC   nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC   polypeptide or its fragment whose expression is inhibited by the
CC   antisense nucleic acid; (4) an antibody capable of specifically binding
CC   the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC   proliferation or the activity of a gene in an operon required for
CC   proliferation; (7) identifying a compound that influences the activity of
CC   the gene product or that has an activity against a biological pathway
CC   required for proliferation, or that inhibits cellular proliferation; (8)
CC   identifying a gene required for cellular proliferation or the biological
CC   pathway in which a proliferation-required gene or its gene product lies
CC   or a gene on which the test compound that inhibits proliferation of an
CC   organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC   compound's activity; (11) a culture comprising strains in which the gene
CC   product is overexpressed or underexpressed; (12) determining the extent
CC   to which each of the strains is present in a culture or collection of
CC   strains; or (13) identifying the target of a compound that inhibits the
CC   proliferation of an organism. The antisense nucleic acids are useful for
CC   identifying proteins or screening for homologous nucleic acids required
CC   for cellular proliferation to isolate candidate molecules for rational
CC   drug discovery programs, or for screening homologous nucleic acids
CC   required for proliferation in cells other than S. aureus, S. typhimurium,
CC   K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC   the target prokaryotic essential genes. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 2234 AA;

Query Match             4.3%;  Score 386; DB 6;  Length 2234;
Best Local Similarity  18.7%;  Pred. No. 7.2e-17;
Matches  339;   Conservative 236;  Mismatches 633;  Indels 608;  Gaps  76;

Qy        21 FTQANNFTSAVSGGVDPR----------TGLYNIQITLGHIVG-------NGNL---------  57
             ::::  |||| :  |||          |:  ::  ||:: |          |||
Db       748 YSKVGHAESBEISHEFVPRRSQDSGFIGMFDYWASIPVLNGKVNATNGNFIMSEKDITLSG 807

Qy        58 -GPTLPLTLSYSPLNKTDIGFGIGFNFGLSV--YDRKNSLLSLS-------------TGEN- 102
              ||| :||||| || ::||| :|| ||| :    |:|| :|:|              ||:
Db       808 RGPDVSVERTYNSQSKKVGLFGTGWSSGLEERVWADGNGNLLLISTDGANIFTRTGDNK 867

Qy       103 ---------YKVIETDKTVKLQQKKLDNLRFEKDLKENCYRIIHKSGDIEVL 145
                      ::|:||
Db       868 YQAPTGIYLEIKQVSGGYEIKDKDQTV----------TFYXSGDAQGR 905

Qy       146 TGFNNNAFDLKVP----------KGLLNPAGHAIYIDWNFEATQPRLNRIYDDLDGHDIPL 196
             :: :   |||             ::  |: :| ::
Db       906 IEYTKDKYGNTTTYEYDGASRLSKVKNASGKELVL----QYDGNNKKAARVI----GPDNKT 959

Qy       197 LNLBEYQGLIKTILFPGQ--KEGYRTELRFLNRQLNSIHNFSLGNENPLTWSFGY------- 250
             |:|       ||:   |    :  ::|:|||
Db       960 ITFNYDGDLLVSSTTPEGKVYKYGYD------NGVLTSIYDPQHTDAKPYKTSYAYENDR 1013

Qy       251 -----TPIGKNGII----GQWITSMTAPGGLKETVNYSNNNQGHHFPQSANLPVLPVVTL 301
                  ||:|||||    :||
Db      1014 LVKVTDPLGKATTLAYNTGSKEVTLTNPKGRKTVYTYND---------AGNPV------- 1057

Qy       302 MKQVPGAGQOPAIQAEYSYTSHNYVGG---------GSNGIWNNKLDNLYGLMTEYN---- 348
              ::|:   :   | |:|                 ||
Db      1058 -KTVEDVGRLNLTTSYEYNANNLVKTTTPKNQTETATYDGNGNVTSVTDEMGTEKFEYNK 1116
```

```
Qy       349 -YGSTESRRYKDKE-------GHDQIVRIERTYNYHLLTSECKQQNGYIQTTETAYYAI  400
              ::  ||:    ::     :::  ::  ::::
Db      1117 DNGIIKATDNEDRKTTVAYVGANTEXSVQTDQGAN----TSSVIHHDQYGNPIETSKELS 1171

Qy       401 IGHNFDSQPSQFOLPKTKTETWRSADNSYRSEITETTFDESGNPLTKVIKDKKTQKIISP  460
             :||:::: |::   |:|  :     :::    :    |    :::
Db      1172 AGGNLIQNPS-FEM--NGTEKWKVDTNNSGSIS--------KDATPAPGGLGGESSLKITTK 1223

Qy       461 STH--WEYYPPAGEVDNCPEPYGFTRFVKKII--------QTPYDSBFKDDP  503
             :::   ::  |:|| ||  ::  ||:  |||:        ||| :: ||||
Db      1224 ATNNDWGYIAAIQEVTLEPNTTYTLSGMVKTDLVNGAAFNVQSLNENGAGIDGGWHDTR 1283

Qy       504 EKFIQYRYSLIGSQSHVTLKIEERHYSATQLLNSTLFQYNTDKSELGRL-LKQTECTKGE  562
             |:     :  :|| || ||: :: :       ::  |:|:|:|:  :::  |||:||||
Db      1284 HNKVQGTSDWNRQ--VTFKTTEQ----TRKVKIYLQVENGGSATSGSAWFDKIQLEKGE 1337

Qy       563 NGKTYSVVHKPTYTKQ--DDTLQQ-SHSITTH-----------DNFTIHRSQVRSRYTGRLFS  611
             |||| :|      :|    ||::|   |::||            |:|  |:|   :|| |
Db      1338 VSSSNPVLNSSFEENWPDGFVPQWVRSCCQHCERNDVSDDSPTGHSSIVMER-------S 1391

Qy       612 DTDTKDIVTQMSYDKLGRLLTRTLNSGTPYANTLTYDYELNNLQDDNRPPFVITTTDVNG  671
             |:    |: :||::  | :  |:|  | |  ::  |:|:|||  |::||| |:||||| :
Db      1392 EYGPNDI----------GYRNRVILNQKAETVTLTAMSKSENVND--------------- 1428

Qy       672 NQLRNEFDGAGRHVSQCLKDSDGDGKFYTIHTQQYDEQGRHHTSTYSDYLTNGRQQTDP-  730
                                                      :: | :|||:|||:|||:
Db      1429 ----------------APDKLSKDYAVLAETYYQDG---TVVNYYTSFPSGTNDW 1464

Qy       731 ----------DKVHLSMSKSYDNWGQIANTHWSYGVSEKITVDPITLTATKLQSNS  777
                      |::: : :  ::|  :     ::||  |:::  |
Db      1465 NRSAAVIPAKKPIQKIBIFLLFRKNNKG------------KVWFDDIRLLEGNALIKNE 1511

Qy       778 NNVQTGKEVTTYTPSQQPIQITLFDEAGHLQSCHTLTRDGWDRVRKETDAIGQCTIYQYD  837
             ||||   ||     |         | |     | :|:    |    :|::  | ||:|:|
Db      1512 YD-NDGNVATY----------DEEGQK--NTFTYDASGNKKSETDEKGNTKLYDYN 1555

Qy       838 NYNRVIQITLPDGTIVNRKYAPFSTDTLITDIRVNGISLGQQTFDGLSRLTQSDGGRVW  897
             |:|  | :| ||| |||
Db      1556 KDNLLTKVTLKNGTSVNYRYD--------------------- 1576

Qy       898 AYTYSAGNDQCPSTVITPDGQFIHYQYQPELDDAVLQVASNEITQQFSYNPVTGALLKAV 957
                                      |:   |       :::    :|
Db      1577 ---HNGNTBKSVMFGGKTQTHKYEYDVDNKNTVYIDALNRRIEN-TYDENANKIKTKM 1631

Qy       958 AEGQSLTPIYYPSGRLKMENINDMKKMSYLWTLRGLENGYTDLTGTIQKISRDTHGRVTQ 1017
               :::  | : :  | :|  :|  ::| :| |        :: | : :::|::  |::|
Db      1632 PNGSILESVYDTADRVVGEK-------------RNGKDSFT---FERDQNGQVTK 1670

Qy      1018 IKD--SSIKTLNYDDLNRHIGSQVTDLATGHMLTTVEFDGLNREIGRKLCDSSGHTLD 1075
             | |   |: : |: : | :  :   |:     :  |  |:
Db      1671 VKDLVNGVERTKTYDKADR-----VTS-AT--------DSRGGKID 1702

Qy      1076 IQQSWL-------KTQQLANRIVKLNGVLQRTEQSYDSRNRLNQ--YKCDGAECPTDKY 1126
             :||          |  |::|   :|  ||::| |:: :|:
Db      1703 ----WAYHDKANSKTEKLKEQTVTQGGY--TNKVSYD-YNTLDQNIRVTDGSQ 1748

Qy      1127 GHSIVTQNFTYDIYGNITACHTTFADGTEDHATFKF--ANP-----TDPCQLTEVHHT 1177
              :: | : :||  :||:  |: :: |:| ||  :|    |||:
Db      1749 -----TYRFDYDGQGNV---RTYTAGNSGSTFNYDQANKIKDLVVGTSNSILLSERYE 1799

Qy      1178 HPDMPDNIRLKYDKAG--------RVINITDNHGNTENFTYDTLGR----LQ 1217
             : :|::|: |: ::|          ::|  : :|||:| :  ||:
Db      1800 YDQGNRTKIKHEGAGGKVTETNFVYDPINQLLNEVLPNGTTKSYTYDGFGFGNRTSVKVIE 1859

Qy      1218 NGQGS---VYGYDPLNRLVSQKTDTLDCEL------------YKCDGAECPTDKY 1244
             |:||:                  ::  :::|
Db      1860 NGKETKSIAATFNEGNQLVKFGNESLTYDVNGNRTSDGKYKYTWNEDDQIVAITKQGENN 1919

Qy      1245 ----YYRETMLVNEVRNGEMLRLRTGETI-----------IAQQ 1274
                 :|::|||: |:  :||:: :||:||:             ::
Db      1920 AFATYKYDEDNRRIEKNVNGQVTRYFYDGDSINPLYETDGNGTVLRQYVSADGARLAMK 1979

Qy      1275 RASKVLLTGTDSQQSVI--LTSDKQNLSQEAYSAYGK-----HKSTANDASILGYNGERA 1327
```

```
Db  467 YPPAGEVDNCPPBYGFTRFVKKIIQTPYDSEFKDDPEKFIQYRYSLIGSQSHVTLKIEE 526
Qy  273 ----------PRPLSDSAFPDTLPGTEY-----------GPDRGIR- 297
        :||                                   : :
Db  527 RHYSATQLLNSTLFQYNTDKSSELGRLLKQTECTKGENGKTYSV---------VHKFTYTQD 579
Qy  298 --LSAWLTHDPAYPESLPGAPLARY----------TYTEAGELLAVYDRSNTQVRAFTYDAQH 349
        :                                  :|:
Db  580 DTLQQSHSITTHDNFTIHRSQVRSRYTGRLFSDTDTKDIVTQMSYDKLGRLLTRTLNSGT 639
Qy  350 PGRMVAH----------------------RYAGR------PEMRYRYDDTGRVVEQLNPAGL 383
        :|:|                          |||
Db  640 PYANTLTYDYELNNLQDDNRPPFVITTD-VNGNQLRNEFDCGAGRHVSQCLKDSDGDGKF 698
Qy  384 SY------QDR-------ITVTDSLNRREVLHTEGGAG--LKRVVKKELADG-- 425
        |||
Db  699 YTIHTQQYDEQRHHTSYSYDYLTNGR--QQTDPDKVHLSMSKSYDNWGQIANTHWSYGV 756
Qy  426 ------------------SVTHSGYDAAGRLTAQTD-- 453
                          :|:
Db  757 SEKITVDPITLTATKQLQSNSNNVQTGKEVTYTPSQQPIQITLFDEAGHLQS----CHT 812
Qy  454 N--VVSGDITDITTPDGRETKFYYNDGNQLTAVVSPDGLESRRAYDEPGRLVSETSRCGD 511
        :|
Db  813 LTRDGWDRVRKE------TDAIGQCTIYQYDNYNRVIQITLPDGTIVNRKYAPFSTDTLIT 867
Qy  512 VIRYAYDNPHSELPATTTDATGSTRQMTWSRYGQLLAFTDCGSYQTRYEYDRFGQMTAVH 571
        :|:
Db  868 DIRVNGISLGQQTFDGLSRLTQSQDG-GRVWAYTYSAGNDQCPSTVITPDGQFIHYQYQP 926
Qy  572 --REEGIS-RYRRYDNRGRLTSVKDAQGHETRYEYNAAGDL--TAVITPDGNRSETQY- 624
        :|||
Db  927 ELDDA---VLQVASNEITQQFSYNPVTGALLKAVAEGQSLTPIYYPSGRLKMENINDMKK 983
Qy  625 ---DAWGKAVSTTQGGLTRSMEYD-LAGRITTLTNENGSRSEFTYDA- 667
        :|:
Db  984 MSYLWTLRGLENGYTDLTGTGIQKISRDTHGRVTQIKDSSIKTTLNVDDLNRHIG------ 1037
Qy  668 ------GRTQRYHYDLTGKLTQSEDEGLVTLMWHYDESDRLTHRTVNGE 720
        :|:
Db  1038 ------SQVTDLATGHMLTTTVEFDGLNRBIGRKLCDSSGHTLDI----QQS 1079
Qy  721 PAEQWQYDEHGWLTEISHLSEBGHQVAVHVGYDDKGRLAGERQTVHNPETGELLMQHETEH 780
        :|:
Db  1080 WLKTQQLANRIVKLNGVLQRTEQYSYDS------RNRL------N 1112
Qy  781 AYNEQGLANRVTP--DSLPRVEWLTYGSSGYLAGMKLGGTPLVEFTRDRLHRETVRSFGNN 838
        :||
Db  1113 QYKCDGAECPTDKYGH------------SIV------TQNFTYDIYGNIT 1144
Qy  839 AYELTSTYTPA--GHLQSQRLNSQVYDRDYDWNDNGDLVRISGPRQTWEYGSATGRLE 895
        :|:
Db  1145 ACHTTFADGTEDHATFKPANPTDPC--QLTEVHHTHPD----MPDNIRL------KY 1189
Qy  896 SVRTLASD------LDIRIPYATDPAGNRLPD-PELHPDSTLTAWPDN-RIAEDAHVYRH 948
        :|:
Db  1190 DKAGRVINITD------NHGNTENFTYDTLGRL--------QNGQGSV---YGYDPL- 1229
Qy  949 DEYGRLTEKTDRIPAGVIRTDDERTHHYHYDSQHRLVFVFYTRIQHGEPLVESRYLYDPLGR 1008
        :|:
Db  1230 ------------------------NRLVSQKTDTLDCELYYR---- 1247
Qy  1009 RMAKRVWRRERDLTGWMSLSRKPEVTWYGWDGDRLTTVQTDTTRIQTVYEPGSFTPLIRV 1068
        :|:
Db  1248 ------------ETMLVNEVRNG---------EMIRLLRTGETIIAQQRAS---- 1277
Qy  1069 ETENGEREKAQRRSLAETLQQEGSENGHGVVFPAELVRLLDRLEEBIRADRVSSESRAWL 1128
        :|:
Db  1278 ------------KVLLTGTDSQ--QSVILTSDKQNLSQEAYSAYGKHKST 1313
Qy  1129 AQCGLTVEQLARQVEPEYTPARKVHFYHCDHRGLPLALLISEDGNTAWRGEYDEWGNQLNE 1188
        :|:
Db  1314 ANDASI---LGYNGERADPVSGVTHLGNGYRSYDPTLMRFHTPDSLSPFG-AGGINPYSY 1369
```

```
Db  1980 AQGQTLJYYHYNPRGDVVAMTNQDKEVVATYEYDAWGNVLTSDTKGIAAD-NPFGYAGYMY 2038
Qy  1328 DPVSGVTHLGNGYRSYDPTLMRFHTPD-----SLSPFGAG-----GINPYSYCLGDPINR 1377
        :|:
Db  2039 DKEIGMYYL----------IARYYNPEHGVFLSVDPDPGDEDDPVTMNGYTYADNNPVMM 2088
Qy  1378 SDPSGHLSWQAWTGIGMGIAGLLLTIATGGMAIAAAGGIAAAIASTSTTALAFGALSVTS 1437
        |||:
Db  2089 TDPDGKWAW----------LVPVVIAGAMVAARFGAKYAI---RYGAKYGKKAVKS 2131
Qy  1438 DITSIVSGALEDASPKASSILGWVSMGMGAAGLAESAIKGGTKLATHLGAFABDGENALL 1497
        :|:
Db  2132 ------------GWDY----GKKVAKSGWNKGKSIAQKIPRIHKVGR---I 2163
Qy  1498 KSTSESSRIKWGVTRSLDREIVRNEEGQVIKDHSRGYTDNFMGKGEQAILVHGDKDGFLY 1557
        :|:
Db  2164 KGDNDKGKGYWGVIYSTTK----------KTGKRTYS---------SFEFHTPHNGHGY 2203
Qy  1558 HTEGNKHNG-KGPYTR 1572
        |||:|:
Db  2204 HLQKNYSKYQGKWNR 2219
```

RESULT 11

```
ADC01365
ID  ADC01365  standard; protein; 1400 AA.
XX
AC  ADC01365;
XX
DT  04-DEC-2003  (first entry)
XX
DE  Enterohaemorragic E. coli 0157:H7-specific protein SEQ ID NO: 1410.
XX
KW  enterohaemorragic; anti-bacterial.
XX
OS  Escherichia coli; 0157:H7.
XX
PN  JP2002355074-A.
XX
PD  10-DEC-2002.
XX
PF  24-JAN-2002; 2002JP-00015959.
XX
PR  24-JAN-2001; 2001JP-00112010.
XX
PA  (UYTS-) UNIV TSUKUBA.
XX
PT  WPI; 2003-451640/43.
XX
PT  Enterohemorragic Escherichia coli 0157:H7-specific nucleic acid molecule
PT  and a polypeptide and its use, a vector and a host cell.
XX
PS  Claim 3; SEQ ID NO 1410; 2067pp; Japanese.
XX
CC  The invention relates to a novel enterohaemorragic Escherichia coli
CC  0157:H7-specific nucleic acid molecule. A polymucleotide of the invention
CC  has anti-bacterial activity. The polypeptide can be used in detection
CC  and/or treatment of 0157:H7 infection. The nucleotide sequence of the
CC  genome of Enterohaemorragic E coli 0157:H7 was determined. The present
CC  sequence represents an E. coli 0157:H7-specific polypeptide of the
CC  invention.
XX
SQ  Sequence 1400 AA;
```

```
Query Match       4.1%;    Score 365;  DB 7;  Length 1400;
Best Local Similarity  20.9%;  Pred. No. 1e-15;
Matches  284;  Conservative 154;  Mismatches 467;  Indels 467;  Gaps 65;
```

```
Qy  418 KTETW-RSADNSYRSEITETTFDESGNPLTKVI---------KDKKTQKIISPSTHWEY 466
        |||:|:|:
Db  218 RTLTYRREAAGDLAGEITGVT-DGAGREFRLVLTTQAQRAEEARKQHTASLSSPDT---- 272
```

```
Qy   192  HDIPLNLEYQGLIKTILTLFPGQKEGYRTELRFLNRQLSIHN---------------F  236
                 :    ::     ||||  :||||| |||||
Db   308  ---KLTGISY----------------TDGEQLLTYAGRQLQSVTDTRGRRLLFAYQADRIT  350

Qy   237  SLGNENPLTWSFGYTPIGKNGILGQWITSMTAPGGLKETV--NYSNNNQCHHFPQS---  290
                :|||: :: : :     |||||  |||   ||| :
Db   351  QVGLPDGMVLAYGYDSQARL----RSVTLQTAAGAVASAIAGYDYGN----ARFPDALTT  402

Qy   291  ------ANLPVLPVTLMKQVPGAGQPAI-QAEYSY-----TSHNYVGG-----GSN  330
                 : :  ::   :      : :     ||| ||
Db   403  HRDEQDQVYASWTVDAQQRVVRSVHGDPTGKIDEATIAYSGNTSTVSNALGNAVTRTGIS  462

Qy   331  GIWNNKLDNLYGLMTEYNYGSTESRR------YKDKEGHDQIVRIERTYNNYHLLTSECKQ  385
                :|:    |: : : :      |||  :||       ||| |
Db   463  KLGQAKVTAVQGLCPECTVGAFKSRSYDGNGYPDQEVDFAGVATDWRYNNRGLLASKIEA  522

Qy   386  QN---GYIQTTETAYYAIIGHNFDSQPSQPQLPKTKTETWRSAD-----NSYRSE  432
                      ||:    |   |||:|     :    ::
Db   523  DNSSGGQKRTLQT---------DWHPS-FRVP-TDQRTYDASDILVARTSWTNSRGQA  570

Qy   433  ITETTFDESGNPLTKVIDKDKKTQKIISPSTHWEYYPPAGEVDNCPPEPYGFTRFVKKIIQ  492
                :|::  :||   |:   |    : :|| | |  || :       ::  ::
Db   571  LTVSRTDPSGGP-TRV---TTQRYCEDSD-----TAAG---NC-  601

Qy   493  TPYDSBFKDDPEKFIQYRYSLIGSQSHVTLKIEERHYSATQLLNSTLFQYNTDKSELGRL  552
                :||:  :         |    : :|   : ::|
Db   602  ------------SLPG---LLLASDGARTDIADSTSYTY---  627

Qy   553  LKQTECT----------KGENGKTYSVVHKFT-YTKQDD-------TLQQSHSITTHDNFTIH  597
                :|| :         ||  | :: :|| :||        ||| | | ||:  |
Db   628  ADDASCTAGASTCPHRKGDRWKVTNALGQTTEYLAYDGAGRPLSIKDTNGIVT--DYTYH  685

Qy   598  -----RSQVRSRYTGRLFSDTDTKDIVTQMSYDKLGRLLTRTLNSGTPYANTLTYDYEL  651
                      |        :  ||| | :||  | : :    :|  |
Db   686  PRGWLTASKVRG------ADASSEADDRITRIDIWPTG--LVRQVTQPDGAFTAFTYD-AA  737

Qy   652  NNLQDDNRPPFVITTTDVNGNQLRNEFDGAGRHVSQCLKDSDGDGKFYTIHTQQYDEQGR  711
                 ||  | :||  ||: | |: :| |   : |  || ::
Db   738  HRLITD---------ITDNAGNTVHYTLDNAGNRVKEDTKDAAATLK--RTLSRVYNQLGQ  786

Qy   712  HHT-----STYSD-----YLTNGRQQTDPDKVHLSMSKSYDNWGQI---------ANTHWS  753
                ||       | |  : ||||| | |   :  | ::|  :: :
Db   787  LKTQATAAASDPTDFAYDANGNATKVTDALATATQSEYDINRLSHTLQDVAGIKADTKFA  846

Qy   754  YGVSEKIT---VDPITLTATKQLQSNSNNVQ-------TGKEVTTYYTPSQQPIQITLFDEAGH  806
                ||| :     :|   |||| |  :|||        |||   |||  | :|  |   |||
Db   847  YDALDNLTKVTDPKGLDTTYDYNGFGELVKLTSPDTG--VTSYT------  894

Qy   807  LQSCHTLTRDGWDRVRKETDAIGQCTIYQYDNYNRVIQITLPDGTIVNRKYAPFSTDTLI  866
                ||  |     : :   || |||  |  : :   :|| ||
Db   895  RAT--------QTDARGNTTAYSYDALNRLTKVTYPTS-----SLDVTY  930

Qy   867  T-DIRVNGISLGGQTFDGLSRLTQSQDGGRVWAYTYSAGNDQCPSTVITPDGQPIHYQYQ  925
                | :|  | | |  ||   ::   |  |  |||:   : :  : |   |  :|  :
Db   931  TYDVTQTACTSG-ETF-SIGRLTKMQDGGAITQCYNRFGD-LVRKVQTSNGTALVLRYD  987

Qy   926  PEL----DDAVLQVASNE--ITQQFSYNPVTGALLKAVAEGQSLTPIYYPSGRL  973
                 ||    |:| | ::   :|   :|  | :||   |:        |  |  |:|
Db   988  YTVGGQLRRMTYPDGAVVDYVRNAQGQTTQVGVTPAGGS--RQVLLGNA--TYYPFGPA  1042

Qy   974  KMENINDMKKMSYLWTL---------RGLENGYTDLTGTIQKISRDTHGRVTQIKDS  1021
                | :  |   | :          |  |:|| ::|  |:|  :  :| |:
Db   1043  AGWTYGNGRTLARQYDLLDYRPQAIQDTRPGGLLDVGF---------GFDPAGNLTALTPA  1092

Qy   1022  SIKTT----LNYDDLNRHIGSQVTDLATGHMLTTTVEFDGLNRREIGRKLCDSSGHTLDIQQ  1078
                 :      |    |||:||  :| ||||:    ::   |  :||  || | | | |  :
Db   1093  GNTTPEIGLGYDALGRLTG--LTDGVTGTVI-------DG--  1123

Qy   1079  SWLKTQQLANRIVKLNGVLQRTEQYSYDSR-NRLNQYKCDGAECPTDKYGHSIVTQNFTY  1137
                       :  :        ::| ::|:|
Db   1124  ------------YSYDATGNRL-------STKVG-TATQIYTY  1146

Qy   1138  DIYGNITACHTTFADGTEDHATFKFANPTDPCQLTEVHHTHPDMPDNIRLKYDKAGRVIN  1197
                 :: :    : : :|   :| :|:   :| |   :|  || :   |  ::|  |||::|
Db   1189  ENPYYILHQPYRLPGGQHDEESGLYYNRNRY--YDPLQGRYITQD---PIGLAGGWNLXNY  1243

Qy   1370  CLGDPINRSDPSG-----HLSWQAWTGIGMGIAGLLLT-----IATG--GMAIAAAGGI  1416
                ||| |  | |:||      : :| | || | | | |        |:|   |  |||| |
Db   1244  PL-NPIIRMDPLGLYNLYQLLYDVWHDDSYGTSSIDITGSGDLISLGGHAGLGVAFAKKK  1302

Qy   1417  AAAIASTSTTALAFGALSVTSDITSIVSGALEDASPKA--SSILGWVSMGMGAAGLAESA  1474
                || |::|  |:||| || :|:|| |:|: | |:||    :  || || :| || |:|:|
Db   1303  GEMLSDICIYATACGHAGIGGGINAAITYSVSKSLPTSGVSNSVG-VTVGGGVGG-----  1356

Qy   1475  IKGGTKLATHLG-AFAEDGENALLKSTSESSRIKWGVTRSL  1514
                 :||| |: | | |  :| |:|  | | :::   | | |:|
Db   1357  ------HFAYTVVDVDNP--ESSTESVGIGAGVDASV  1386


RESULT 12
ADG30698
ID   ADG30698 standard; protein; 1510 AA.
XX
AC   ADG30698;
XX
DT   26-FEB-2004  (first entry)
XX
DE   Xanthomonas axonopodis pv citri plant pathology-related XAC3245 protein.
XX
KW   Xanthomonas microorganism; plant; pathology; bacterial pest; Xac; Xcc;
KW   XAC.
XX
OS   Xanthomonas axonopodis pv. citri.
XX
PN   WO2003089647-A1.
XX
PD   30-OCT-2003.
XX
PF   22-APR-2003; 2003WO-BR000060.
XX
PR   22-APR-2002; 2002US-0374620P.
XX
PA   (AMPA-) FUNDACAO AMPARO A PESQUISA DO ESTADO.
XX
PI   Da Silva ACR, Farah SC, Quaggio RB, Reinach FDC, Ferro JA;
PI   De Oliveira JCF, De Laia ML, Setubal JC, Furlan LR;
XX
DR   WPI; 2003-865444/80.
DR   N-PSDB; ADG30697.
XX
PT   New nucleic acid molecule from a Xanthomonas microorganism, useful in
PT   determining the presence of Xanthomonas bacteria in a sample.
XX
PS   Claim 8; SEQ ID NO 55; 145pp; English.
XX
CC   The invention relates to a novel isolated nucleic acid molecule from a
CC   Xanthomonas microorganism where the nucleic acid molecule is associated
CC   with pathogenicity caused by the Xanthomonas microorganism, or its
CC   variant, that causes reduced or enhanced pathogenicity. The nucleic acid
CC   of the invention may be useful in detecting the presence of Xanthomonas
CC   bacteria in a sample, as well as in plant pathology, for example, for
CC   identifying nucleic acid molecules and proteins involved in pathology
CC   caused by bacterial pests. The current sequence is that of the
CC   Xanthomonas axonopodis pv. citri (Xac) plant pathology-related XAC
CC   protein of the invention.
XX
SQ   Sequence 1510 AA;

Query Match             4.1%;  Score 361;  DB 7;  Length 1510;
Best Local Similarity  21.2%;  Pred. No. 2.2e-15;
Matches  304;  Conservative 158;  Mismatches 466;  Indels 504;  Gaps 75;

Qy   144  VLTGFNNNAFDLK-------VPKKLLNPAGHAIYIDWNFEATQPRLNRIYDDLDG  191
                :: : |        :|:  ||   : |  ::|  :  ||:|||
Db   250  ILTRPNGNFYTFKKSANGTWTPDWDVRETLSEVQDNGTLTGWQVTAIDDSKEQF--DLDG  307
```

```
CC      identifying a gene required for cellular proliferation or the biological
CC      pathway in which a proliferation-required gene or its gene product lies
CC      or a gene on which the test compound that inhibits proliferation of an
CC      organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC      compound's activity; (11) a culture comprising strains in which the gene
CC      product is overexpressed or underexpressed; (12) determining the extent
CC      to which each of the strains is present in a culture or collection of
CC      strains; or (13) identifying the target of a compound that inhibits the
CC      proliferation of an organism. The antisense nucleic acids are useful for
CC      identifying proteins or screening for homologous nucleic acids required
CC      for cellular proliferation to isolate candidate molecules for rational
CC      drug discovery programs, or for screening homologous nucleic acids
CC      required for proliferation in cells other than S. aureus, S. typhimurium,
CC      K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC      the target prokaryotic essential genes. Note: The sequence data for this
CC      patent did not form part of the printed specification, but was obtained
CC      in electronic format directly from WIPO at
CC      ftp.wipo.int/pub/published_pct_sequences
XX
SQ      Sequence 843 AA;

Query Match             4.0%;  Score 353.5; DB 6;  Length 843;
Best Local Similarity   22.8%;  Pred. No. 2.8e-15;
Matches  223;  Conservative 121;  Mismatches 308;  Indels 327;  Gaps  48;

Qy    542 YNTDKSELGR----LLKQTECTKGENGKTYSVVHKFYTYKQDDTLQQSHSITTHDNFTIH 597
          :                ::|| |:  |||||:||:        ||::  : :
Db     32 HHTDSGEYRLDWNLAERSLCVTDSMGRT------------------RCHWWDAQGLVTAY 74

Qy    598 RSQVRSRYTGR-----LFSDTDTKDIVTQMSYDKLGRLLTRTLNSGTPYANTLTYDYE 650
          :::||||:||||    ||:|||     :|:| : ||:|:| |||:|:||:| |||  ||
Db     75 RDEAGQMTTFRWSDEERLLLGMTDAQGGKWRVYDRLGHL---------TETHD-P 120

Qy    651 LNNLQDDNRPPF--VITTTDVNGNQLRNEFD-----------GAGRHVSQCL 689
          |                |:                     ||||   ||
Db    121 LGRVEQTQWHFVWHQPETEVDAAGVAWRYEYDERGNLQAVSDPLHQRTVYGYDRH-GQVV 179

Qy    690 K--DSDGDKFYTIHTQQYDBQG---RH------HTSTYSDYLTNGRQQTDPDKVHLSMS 738
          |               : |:|        ::|:| |:||     ||
Db    180 RITDARGGDKYL-----QWNEDGQLMRHTDCSGSQTAWFYDERTRLERVTDAES--NSTR 232

Qy    739 KSYDNWGQIANTHWSVGVSEKITVDPITLATKQLQSNSNNVQTGKEVTTVTPSQQPIQI 798
          |: || |   : : : ::                     |:| |  :|||: : |::|||
Db    233 YSYDGNGHLTBVMFADGRTERYQPD------------AAGRLVKYTSPAG---QI 272

Qy    799 TLFDEAGHLQSCHTLTRDGWDRVRKETDAIGQCTIYQYDNYNRVIQITLPDGTIVNRKYA 858
          ||   ||     |                  |:| |: :||::|  :
Db    273 TRWQ-----RDGQGRVRRQTDATGRRTAYEYDAYGRLTTLTNENGESYRPRY- 319

Qy    859 PFSTDTLITDIRVNGISLGQQTFDGLSRLTQSQD-GGRVWAYTYSAGNDQCP-------- 909
              ::|  ||  :|||   :| | | |      ||||  |:|:||
Db    320 ------------------------DVLDRVTEQTDPGGSRRAYGYNALNAVTAVIYGGERG 356

Qy    910 ----------------STVITPDGQFIHYQYQPELDDAVLQV------ASNEITQQF 944
                          ||:||| |||||| |||:| |:|:             |::|:||
Db    357 GEIRHGLERDAAGRLTAKITPE--TRTEYRYDAADRLLEIRRRRHDAAEGGEPEVI-RF 412

Qy    945 SYNPVTGALLKAVAEG-----------QSLTPIYPSGRLKMENINDMKKMSYLWTLRGLEN 995
          |: |  | |||:|      |:|:|||: :|:||  ::| | :|  |  |
Db    413 SYDSAGNLLSSETAQGVLQHRYDVQGNRTETQMPDGR------------------TLRYLYY 456

Qy    996 GYTDLTGTIQKISRDTHGRVTQIKDSSIKTLNYDLNRHIGSQVTDLATGHMLTTTVEF 1055
          |:||:||:| |||   ||||: || : :|  ||| ||  |:||:| |
Db    457 G----SGHLQQI---------------------------------NGRDVISEFTR---- 476

Qy   1056 DGLNRREIGRKLCDSSGHTLDIQQSWLKTQQLANRIV--KLNGVLQRT--EQYSYDSRNR 1110
          |  |||:||||  |::| ||| ||:|:| ||| | :|   |||:|:|    ||||||::|
Db    477 DHLHREVQR-----SQGR-LDMRRMYDRTGRLTRKLTCKGMRGVVPETFIDREYAYSGQDE 531

Qy   1111 LNQYKCDGAECPTDKYGHSIVTQNFTYDIYGNITAC-HTTFADGTEDHATFKFANPTDPC 1169
          | :           :| ||:|:|:|:|  |:  :||   ::|
Db    532 LLK------KRHSRQGVTDYF-YDTTGRITACRNEAYLDSWQYDAA---ANLLDRR 577
```

```
Qy  1170 Q---------------LTEVHHTHPDMPDNIRLKYDKAGRVINITDNHGNTENFTYDTL 1213
              :|    |||:  :|    :||:   :|
Db   578 QGETAQAGAGSVVPFNRITSYRGLH-------YRYDEYGRVVEKRGRNG-TQHYRWDAB 628

Qy  1214 GRL-------QNGQGSVYG--YDPLNRLVSQKTDTLDCELYRETMLVNEVRNGEMIRLLR 1265
              |||          :|   ||        |  |    |:   |
Db   629 HRLTEVAVTRGGTVRRYGYVYDAPGRRVEKHELDAEGKPYNRTTFIWDGMRLAQECRLGR 688

Qy  1266 TGETIAQQR-----------ASKVLLTGTD-SQQSVILTSDKQNLSQEA-YSAY 1307
              :|  |             ||:    |   || |     :|
Db   689 SSSLYIYSDRGSHEPLARVDRAAPGEADEVLYYHTDVNGAPEEMTDGGNIVWEAGYQVW 748

Qy  1308 GK---HKSTANDASILGYNGERADPVSGVTHLGNGYRSYDPTLMRFHTPDSLSPFG-AGG 1363
              ||     :    |:|    || |:   |||  |||   |
Db   749 GNLTHEKETRPVQQNLRFQGGYLDRETGLHY--NLYRFYDPDIGKFISGD---PIGLAGG 803

Qy  1364 INPYSSYCLGDPINRSDPSG 1382
              ||:|     ||:   |
Db   804 INLYQYA-FNPLSYIDPLG 821
```

RESULT 14
ABU19676
ID   ABU19676 standard; protein; 1515 AA.
XX
AC   ABU19676;
XX
DT   19-JUN-2003 (first entry)
XX
DE   Protein encoded by Prokaryotic essential gene #5203.
XX
KW   Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS   Borrelia cepacia.
XX
PN   WO200277183-A2.
XX
PD   03-OCT-2002.
XX
PF   21-MAR-2002; 2002WO-US009107.
XX
PR   21-MAR-2001; 2001US-00815242.
PR   06-SEP-2001; 2001US-00948993.
PR   25-OCT-2001; 2001US-0342923P.
PR   08-FEB-2002; 2002US-00072851.
PR   06-MAR-2002; 2002US-0362699P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI   Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR   WPI; 2003-029926/02.
DR   N-PSDB; ACA23546.
XX
PT   New antisense nucleic acids, useful for identifying proteins or screening
PT   for homologous nucleic acids required for cellular proliferation to
PT   isolate candidate molecules for rational drug discovery programs.
XX
PS   Claim 25; SEQ ID NO 47600; 1766pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising any one of
CC   the 6213 antisense sequences given in the specification where expression
CC   of the nucleic acid inhibits proliferation of a cell. Also included are:
CC   (1) a vector comprising a promoter operably linked to the nucleic acid
CC   encoding a polypeptide whose expression is inhibited by the antisense
CC   nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC   polypeptide or its fragment whose expression is inhibited by the
CC   antisense nucleic acid; (4) an antibody capable of specifically binding
CC   the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC   proliferation or the activity of a gene in an operon required for
CC   proliferation;(7) identifying a compound that influences the activity of
CC   the gene product or that has an activity against a biological pathway
CC   required for proliferation, or that inhibits cellular proliferation; (8)
CC   identifying a gene required for cellular proliferation or the biological
CC   pathway in which a proliferation-required gene or its gene product lies
CC   or a gene on which the test compound that inhibits proliferation of an
CC   organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC   compound's activity; (11) a culture comprising strains in which the gene
CC   product is overexpressed or underexpressed; (12) determining the extent
CC   to which each of the strains is present in a culture or collection of
CC   strains; or (13) identifying the target of a compound that inhibits the
CC   proliferation of an organism. The antisense nucleic acids are useful for
CC   identifying proteins or screening for homologous nucleic acids required
CC   for cellular proliferation to isolate candidate molecules for rational
CC   drug discovery programs, or for screening homologous nucleic acids
CC   required for proliferation in cells other than S. aureus, S. typhimurium,
CC   K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC   the target prokaryotic essential genes. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1515 AA;

Query Match   3.9%;   Score 350; DB 6; Length 1515;
Best Local Similarity   22.1%;   Pred. No. 1.3e-14;
Matches 232; Conservative 126; Mismatches 314; Indels 380; Gaps 59;

Qy   481 YGFTRFVKKIIQTPYDSEFKDDPEKFIQYRYSLIGSQSHVTLKIEERHYSATQLLNSTLF 540
             :|| :   ::| ::
Db   543 FQFGQFYYAYDEHGWMTQWRDTDQTDVRYRYDTAG-------RVVE---------TGTRQ 586

Qy   541 QYNTDKSELGRLLKQTECTKGENGKTYSVVHKFTYTKQDDTLQQSHS---ITTHDNFTI 596
             ::| || | :|:| ::
Db   587 GYHT-----GRFIYEAGCT-----RVIDVDGEWTYAYNDEGLVTAETDFLGHCTYSEWEL 636

Qy   597 HRSQVR--------SRY-TGRLFS----------------DTDTKDIVT--------QMSY 624
             :   ::   :
Db   637 GRLMARIDPLGRRTDYRYDERGQLTSVVESSGRTVDFDYDDEQRLTGARLPNGGTIKLEY 696

Qy   625 DKLGRLLTRTLNSGTPYANTLTYDY--------BL----------NNLQDDNRPPFVITTDV-- 669
             |::|| : | |: |
Db   697 DHLSRLIART-----EPDGNKTTYRYGPRGELLRVVQGDRETRLDYDDR---LRLTDIEL 748

Qy   670 -NGNQLRNEFDGAGRHVSQCLKDSDGDGKFYTIHTQQYDEQGRHHTSTYSDYLTNGR---- 725
                ::| || || | |:|
Db   749 PTGARFRRKIDALGR-----LLEETSPDG----HVTRYD----------YADGPANPRGLL 790

Qy   726 -QQTDPDKVHLSMSKSYDNWGQIANTHWSYGVSEKITV---DPITLTATKQLQSNSNNVQ 781
              :   : |::   ::
Db   791 SAVTRPD-------------GSVSRARYN---SBSLPVEWIDPLGRTIQR-------- 824

Qy   782 TGKEVTTVTPSQQPIQITLFDEAGH----------LQSCHTLTRDGWDRVRK 823
               :| || ||
Db   825 ------TYGPFD--LLTASIDAAGHATRPEYDHATRLTKVINALGETYTYRYDAAGRLAA 876

Qy   824 ETDAIGQCTIYQYDNYNRVIQITLPDGTIVNRKYAPFSTDTLITDIRVNGISLGQ--QTF 881
              || || ||||| |:
Db   877 EIDWGGRATEYDRDRDAVGRLLTKTLPDG-----------------GQWRYTY 910

Qy   882 DGLSRLTQSQDGGRVWAYTYSAGNDQCPSTVITPDGQFIHYQYPELDDAVLQVASNEIT 941
             | :| || | :| : || :: |
Db   911 DASDRLIEIDAGDVKLAYRYDASG--------------RLASAEVQGEHTHVT 949

Qy   942 QQFSYNPVTGALLKAVAEGQSLTPIYPYSGRLKMENINDMKKMSYLWTLRGLENGYTDLT 1001
              :|: ||:| || :|::|| | |
Db   950 -RFAYDR-NGRLIGEDQHGBLLRHVYDADGQRRL-RMTPRRETTYAY-------- 993

Qy 1002 GTIQKISRDTHGRVTQIKDSSIKTLNYDDLNRHIGSQVTDL-------ATG---HMLT 1050
             |:| :| |:| || |||:| | ||
Db   994 ------DVSGALTQVG----QLTIRRDGLGREIGREAGDFVAQQYDALGRIRRQIAG 1041

Qy 1051 TTVBFDGLNREIGRKLCDSSGHTLDIQQSWLKTQQLANRIVKLNGVLQRTEQYSYDSRNR 1110
              :| || | :|
Db 1042 PAVAFDALQADPARAL--------------EQLTRQV-------YHYDAAGQ 1072
```

```
Qy  1111 LNQYKCDGAECPT---DKYGHSIV-------TQNFTYDIYGNITACHTTFADGTEDHATF 1160
              |:|:|       ||            :::|||
Db  1073 LERVD-TGADTLYQRDERGQIICABSLLQPSEHFRYDAVMNI-AAHGQRA---------- 1121

Qy  1161 KFANPTDPCQLTEVHHTH---PDMPDNIRLKYDKAGRVINIT-DNHG--NTENFTYDT 1212
              ||:||       |:| ||||:|||                :|:|
Db  1122 --PVD-----AHHYRRGGLPEQVGVARYKYDARGRTIEKTVEQPGVRPKTWQYTWDG 1171

Qy  1213 LGRL----QNGGQSVVGYDPLNRLVSQKTDTLDCELYYRET--------MLVN---EV 1255
              |||              ||:|:      :          |:
Db  1172 LNRLVRVVTPERGVWA-YRTDAFNRRIEKQ------QVGGRETVKFLWDGPMLAERWIEQ 1224

Qy  1256 RNGEMIRLLRTGETI----------IAQQRASKV--LLTGTDSQQSVILTSDKQNLSQE 1302
              |:|              |||      :||||
Db  1225 RDG------TTGQVTWHIEPGSFLPLAQETDDGLFPLLTDQIGRPKTVPDEQGRPVWKA 1278

Qy  1303 AYSAYGK---HKSTANDA--------SILGYNGERADPVSGVTHLGNGYRSYDPTLMRFH 1351
              |||:|||         |||  |:::|        |||
Db  1279 AYSLWGKLLPVKRPANDADCGATSIDTTLRFSGQWADDETGLNYNLNRY--YDPDSGQYL 1336

Qy  1352 TPDSLSPFG-AGINPYSYCLGDPINRSDPSG 1382
              :|    |     |:|
Db  1337 SAD---PIGLLGGARTQAY-VHDPSQWIDPLG 1364


RESULT 15
ABU15135
ID   ABU15135 standard; protein; 1397 AA.
XX
AC   ABU15135;
XX
DT   19-JUN-2003  (first entry)
XX
DE   Protein encoded by Prokaryotic essential gene #662.
XX
KW   Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS   Escherichia coli.
XX
PN   WO200277183-A2.
XX
PD   03-OCT-2002.
XX
PF   21-MAR-2002; 2002WO-US009107.
XX
PR   21-MAR-2001; 2001US-00815242.
PR   06-SEP-2001; 2001US-00948993.
PR   25-OCT-2001; 2001US-0342923P.
PR   08-FEB-2002; 2002US-00072851.
PR   06-MAR-2002; 2002US-0362699P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Wang L,    Zamudio C,    Malone C,    Haselbeck R,    Ohlsen KL,    ZysKind JW;
PI   Wall D,    Trawick JD,    Carr GJ,    Yamamoto R,    Forsyth RA,    Xu HH;
XX
DR   WPI; 2003-029926/02.
DR   N-PSDB; ACA19005.
XX
PT   New antisense nucleic acids, useful for identifying proteins or screening
PT   for homologous nucleic acids required for cellular proliferation to
PT   isolate candidate molecules for rational drug discovery programs.
XX
PS   Claim 25; SEQ ID NO 43059; 1766pp; English.
XX
CC   The invention relates to an isolated nucleic acid comprising any one of
CC   the 6213 antisense sequences given in the specification where expression
CC   of the nucleic acid inhibits proliferation of a cell. Also included are:
CC   (1) a vector comprising a promoter operably linked to the nucleic acid
CC   encoding a polypeptide whose expression is inhibited by the antisense
CC   nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC   polypeptide or its fragment whose expression is inhibited by the
```

```
CC   antisense nucleic acid; (4) an antibody capable of specifically binding
CC   the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC   proliferation or the activity of a gene in an operon required for
CC   proliferation; (7) identifying a compound that influences the activity of
CC   the gene product or that has an activity against a biological pathway
CC   required for proliferation, or that inhibits cellular proliferation; (8)
CC   identifying a gene required for cellular proliferation or the biological
CC   pathway in which a proliferation-required gene or its gene product lies
CC   or a gene on which the test compound that inhibits proliferation of an
CC   organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC   compound's activity; (11) a culture comprising strains in which the gene
CC   product is overexpressed or underexpressed; (12) determining the extent
CC   to which each of the strains is present in a culture or collection of
CC   strains; or (13) identifying the target of a compound that inhibits the
CC   proliferation of an organism. The antisense nucleic acids are useful for
CC   identifying proteins or screening for homologous nucleic acids required
CC   for cellular proliferation to isolate candidate molecules for rational
CC   drug discovery programs, or for screening homologous nucleic acids
CC   required for proliferation in cells other than S. aureus, S. typhimurium,
CC   K. pneumoniae or P. aeruginosa.The present sequence is encoded by one of
CC   the target prokaryotic essential genes. Note: The sequence data for this
CC   patent did not form part of the printed specification, but was obtained
CC   in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1397 AA;

Query Match             3.9%;   Score 342; DB 6;  Length 1397;
Best Local Similarity  21.4%;   Pred. No. 4.2e-14;
Matches  256;  Conservative 157;  Mismatches 382;  Indels 404;  Gaps 63;

Qy  418 KTETW-RSADNSYRSEITETTFDESGNPLTKVI------KDKKTQKIIS-----PS----- 461
              : ||:                                |||||:||||     |:
Db  218 RTQFTHREAAGEFSGEITGVT-DGAGRHFRLVLTTQAQRAEEARQQAISGGTEPSAFPDT 276

Qy  462 -----------------------THWEYPPAGEVDNCPPEP---YGFT------- RFVK 488
                                     |||:|                 ||:||
Db  277 LPGYTEYGRDNGIRLSAVWLTHDPEYP-----ENLPAAPLVRYGWTPRGELAAVYDRSNT 331

Qy  489 KIIQTPYDSEFK----------DDPEKFIQYRYSLIGSQSHVTLKIEERHYSATQLLNSTL 539
              :: ||:|||           | |||    :||  :
Db  332 QVRSFTYDDKYRGRMVAHRHTGRPE--ICYRYD--SDGRVTEQLNPAGLSYT-------- 379

Qy  540 FQYNTDKSELGRLLKQTEC--TKGENGKTYSVVHKFPTYTKQDDTLQQSHSITTHDNFTIH 597
              |                     ||:||||:::
Db  380 YQYEKDRITITDSLNRREVLHTQGEGG-LKRVVKK-------------EHADGSVT 421

Qy  598 RSQVRSRYTGRLFSDTDTKDIVTQMSYDXLGRLLTRTLNSGTPYANTLTYDYELNNLQDD 657
              : |:|||||               ||   |||  |||
Db  422 QSQFDA--VGRLRAQTDAAGRTTEYSPDVVTGLITRIT---TPDGRASAFYY-------- 468

Qy  658 NRPPFVITTDVNGNQLRNEFDGAGRHVSQCLKDSD------GDGKFYTIHTQQYDEQGRH 712
              | |                      :              |||:|||:||||:|||
Db  469 NHHSQLTSATGDPGLEIRREYDEWGRLIQETAPDGDITRYRYDNPHSDLPCATEDATGSR 528

Qy  713 HTSTYSDYLTNGRQQTDPDKVHLSMSKSYDNWGQIANTHWSYGVSEKITVDPI-TLTATK 771
              |  :||  :      ::: | :|:| :|||  ||:
Db  529 KTMTWSRY---GQLLSFTDCSGYVTRYDHDRFGQVTAVHREEGLSQYRAYDSRGQLIAVK 585

Qy  772 QLQSNSNNVQTGKEVTTYTPSQQPIQITLFDEAGHLQSCHTLTRDGWDRVRKETDAIGQ- 830
              |:  :|   |  || :     :             ||  |||
Db  586 DTQGHETRYE------------------YNAAGDLTT--VIAPDG-SRNGTQYDAWGKA 623

Qy  831 -CTI------YQYDNYRNVIQITLPDGTIVNRKYAPFSTDLITDIRVNGISLGQQTF- 881
               |||          |:|| | ||     ::      |||:||
Db  624 ICTTQGGLTRSMEYDAAGRVIRLTSENGS-------HTTFRYDVLDRLIQETGFDGRTQRYH 678

Qy  882 -DGLSRLTQSQDGGRVWAYTYSAGNDQCPSTVITPDGQFIHYQYQPELDDAVLQVASNEI 940
                 |:: |     :|||  ||:
Db  679 HDLTGKLIRSEDEGLV-----------------------THWHYD-EADRLTHRTVNGET 714

Qy  941 TQQFSYNPVTGAL-LKAVAEGQSLTPIY-YPS-GRLKMENI-----NDM--KKMSY 986
              ::: |:  :  ||| :          :||  |  :||
Db  :
```

```
Db    715  AERWQYDE-RGWLTDISHISEGHRVTVHYGVDSKGRLASEHLTVHHPQTNELLMQHETRH  773

Qy    987  LWTLRGLENGYTDLTGTIQKISRDTHGRVTQIKDSSIKTTLNYDDLNRHIGSQVTDLATG 1046
                :  :||_|       : :: |  _|           ::|: _|
Db    774  AYNAQGLAN--RCIPDSLPAVEWLTYG----------------------SGWLSGMKLG  808

Qy   1047  HMLTTTVEF--DGLNREIGRKLCDSSGHTLDIQQSWLKTQQLANRIVKLNGVLQRTEQYS 1104
                 |:|: |:||   |_|:|       :: ::| || :|  ||_|: |:
Db    809  D--TPLVEYTRDRLHRETLR-----SFGRYELTTAYTPAGQLQSQ--HLNSLLS-DRDYT  858

Qy   1105  YDSRNRLNQYKCDGAECPTDKYGHSIVTQNFTYDIYGNITACHTTFAD------GTEDHA 1158
            ::                    |:::|  ||_||_|      |||||_|:
Db    859  WNDNGELIR-----ISSPRQ------TRSYSYSTTGRLTGVHTTAANLDIRIPYTTDPA  906

Qy   1159  TFKFANPTDPCQLTEVHHTHPD-----MPDN-------IRLKYDKAGRVINITD------ 1200
                 :_|          |||        |||         ||: || :|:
Db    907  GNRLPDP-------ELHPDSALSMWPDNRIARDAHYLYRDRHGRLTEKTDLIPEGV  956

Qy   1201  ---NHGNTENPTYDTIGRLQNGQGSVYG---------YDPLNRLVSQKTDTLDCEL---- 1244
                  :  ||_|:  :: :          ||_|_|::::   :_|
Db    957  IRTDDERTHRYHYDSQHRLVHYTRTQYAEPLVESRYLYDPLGRRVAKRVWRRRDLTGWM 1016

Qy   1245  ---------YY-----RETMLVNE------VRNGEMIRLLR----TGETIIAQQRA--- 1276
                      :|            :         ||_|_|:|_|   |:|:
Db   1017  SLSRKPQVTWYGWDGDRLTTIQNDRTRIQTIYQPGSFTPLIRVETATGELAKTQRRSLAD 1076

Qy   1277  ---------SKVLLTGTDSQQSVLTSDKQNLSQEA-------------------------1303
                     ||:          || : ||_|:|:
Db   1077  TLQQSGGEDGGSVVFPPVLVQMLDRLESEILA---DRVSEESRRWLASCGLTVAQMQSQM 1133

Qy   1304  ------------------------------YSAYGKHKSTANDAS------ILGYNG 1324
                                         : :| : || ::  :: :  :
Db   1134  DPVVTPARKIHLYHCDHRGLPLALISTEGTTAWYAEYDEWGNLLNEENPHQLQQLIRLPG 1193

Qy   1325  ERADPVSGVTHLGNGYRSYDPTLMRFHTPDSLSPFG-AGGINPYSYCLGDPINRSDPSG 1382
             ::  |:  |:    :| : _|||  ::|:  ::|:  |::
Db   1194  QQYDEESGLYY--NRHRYYDPLQGRYITQD---PIGLKGGWNFYQYPL-NPISNIDPLG 1246
```

Search completed: February 9, 2005, 18:17:39
Job time : 224 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 9, 2005, 18:10:24 ; Search time 50 Seconds
(without alignments)
2497.758 Million cell updates/sec

Title: US-09-889-874A-23
Perfect score: 8879
Sequence: 1 VYIKFLKLFRRITMSDNNEF.........PRKIILGRTEKTVKPKTFRP 1673

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Match Length	DB	ID	Description
1	1638	18.4	1584	3	US-09-251-645-6	Sequence 6, Appli
2	339	3.8	1426	4	US-09-492-709A-340	Sequence 340, App
3	331.5	3.7	1377	4	US-09-711-164-467	Sequence 467, App
4	317	3.6	998	4	US-09-252-991A-28424	Sequence 28424, A
5	313	3.5	1627	4	US-09-328-352-6604	Sequence 6604, Ap
6	306	3.4	1586	4	US-09-543-681A-5329	Sequence 5329, Ap
7	294	3.3	1439	4	US-09-543-681A-7560	Sequence 7560, Ap
8	292	3.3	1626	4	US-09-252-991A-23805	Sequence 23805, A
9	285.5	3.2	1596	4	US-09-328-352-5542	Sequence 5542, Ap
10	277.5	3.1	974	4	US-09-252-991A-23640	Sequence 23640, A
11	271.5	3.1	1665	4	US-09-543-681A-4476	Sequence 4476, Ap
12	238.5	2.7	1043	4	US-08-851-567B-61	Sequence 61, Appl
13	235.5	2.7	804	4	US-09-328-352-5545	Sequence 5545, Ap
14	223	2.5	1128	4	US-09-252-991A-31032	Sequence 31032, A
15	222.5	2.5	3290	4	US-09-328-352-5486	Sequence 5486, Ap
16	220.5	2.5	1183	2	US-09-447-031A-2	Sequence 2, Appli
17	218.5	2.5	2123	3	US-08-968-685A-10	Sequence 10, Appl
18	214	2.4	2504	4	US-09-328-352-5821	Sequence 5821, Ap
19	212.5	2.4	954	3	US-09-251-645-12	Sequence 12, Appl
20	211.5	2.4	10182	3	US-09-134-001C-3159	Sequence 3159, Ap
21	207	2.3	2314	4	US-09-268-347-49	Sequence 49, Appl
22	202.5	2.3	2777	4	US-09-543-681A-6124	Sequence 6124, Ap
23	201	2.3	1739	4	US-09-540-236-3739	Sequence 3739, Ap
24	201	2.3	2057	4	US-09-499-203-2	Sequence 2, Appli
25	199.5	2.2	1092	3	US-09-147-405B-15	Sequence 15, Appl
26	198.5	2.2	1004	3	US-09-268-347-30	Sequence 30, Appl
27	197.5	2.2	1726	4	US-09-700-227-2	Sequence 2, Appli
28	194.5	2.2	5024	4	US-09-710-279-2964	Sequence 2964, Ap
29	190	2.1	2385	4	US-09-543-681A-6304	Sequence 6304, Ap
30	188.5	2.1	2736	4	US-09-252-991A-30227	Sequence 30227, A
31	185.5	2.1	1166	4	US-09-200-650E-7	Sequence 7, Appli
32	185.5	2.1	2383	4	US-09-492-709A-302	Sequence 302, App
33	184.5	2.1	2142	4	US-09-540-236-3459	Sequence 3459, Ap
34	184	2.1	3696	3	US-09-134-001C-5080	Sequence 5080, Ap
35	183	2.1	1222	4	US-09-206-942-37	Sequence 37, Appl
36	183	2.1	1228	4	US-09-206-942-34	Sequence 34, Appl
37	182.5	2.1	1565	4	US-08-851-567B-59	Sequence 59, Appl
38	182	2.0	1861	2	US-08-790-912-4	Sequence 4, Appli
39	181.5	2.0	2315	4	US-09-543-681A-5434	Sequence 5434, Ap
40	181	2.0	1228	4	US-09-463-402-2	Sequence 2, Appli
41	181	2.0	1228	4	US-09-889-572-2	Sequence 2, Appli
42	181	2.0	1228	4	US-09-117-447-2	Sequence 2, Appli
43	180.5	2.0	930	3	US-09-134-001C-5314	Sequence 5314, Ap
44	180.5	2.0	930	4	US-09-386-962C-10	Sequence 10, Appl
45	180	2.0	1683	3	US-08-755-587-183	Sequence 183, App

ALIGNMENTS

RESULT 1
US-09-251-645-6

; Sequence 6, Application US/09251645
; Patent No. 6281413
; GENERAL INFORMATION:
; APPLICANT: Kramer, Vance C.
; APPLICANT: Morgan, Michael K.
; APPLICANT: Anderson, Arne R.
; APPLICANT: Hart, Hope
; APPLICANT: Warren, Gregory W.
; APPLICANT: Dunn, Martha
; APPLICANT: Chen, Jeng S.
; TITLE OF INVENTION: NOVEL INSECTICIDAL TOXINS FROM PHOTORHABDUS LUMINESCENS
; TITLE OF INVENTION: AND NUCLEIC ACID SEQUENCES CODING THEREFOR
; FILE REFERENCE: CGC1963/A
; CURRENT APPLICATION NUMBER: US/09/251,645
; CURRENT FILING DATE: 1999-02-17
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 6
; LENGTH: 1584
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-09-251-645-6

Query Match 18.4%; Best Local Similarity 31.0%; Score 1638; DB 3; Length 1584;
Matches 493; Conservative 233; Mismatches 682; Indels 182; Gaps 54;

```
Qy     19 EFFTQANNFTSAVSGGVDPRTGLYNIQITLGHIVGN--GNLGPTLPLTLSYSPLNKTDIG  76
                :   ::   :|  || ::  |  : |:|:||  |||     ||   |:  |::|::|||
Db      5 DIYSNAFNFGSYINTGVDPRTGQYSANINITLRPNNVGNSEQT--LSLSFSPLTTLNNG  62

Qy     77 FGIGFNFGLSVYDRKNSLLSLSSTGENYK---VIETKDTVKLQQKKLDNLRFEKDLKENCY 133
           ||||  |::      |||  :     :      :|  : ::|||: | ||   ||    :
Db     63 FGIGWRFSLTTLDIKTLTFSRANGEQFKCKPLPPNNNDLSFKDKKLKDLRVK-LDSNTF 121

Qy    134 RIIHKSGDIEVLTGFNNNAPDLKVPKKLLNPAGHAIYIDWNFEATQPRLNRIYDDLDGHD 193
             :: |  || :|  ||   :: :|  | : |  |   :  ::|
Db    122 YVYNKNGIIEILKRIGSS--DIAKTVALEFPDGEAFDLIYNSR--------------- 162

Qy    194 IPLLNLEYQGLIKTLITL-FPGQKEGYRTELRFL-NRQLNSIHNFSLGNENPLTWSFGYT 251
              :  ::    ||  |:  |::    | |  |::  |  :::|: |
Db    163 FALSEIKYRVTGKTYLKLNYSGNN---CTSVEYPDDNNISAKIAFDYRNDYLITVTVPYD 219

Qy    252 PIGK------------NGILGQWITSMTAPGGLKETVNYSNNNQGHHFPQSANLPVLPY 298
                             ||   |  :   |:  ::   |    ||    |  | | |
Db    220 ASGPIDSARFKMTYQTLKGVF-PVISTFRIPTGYEVLVSYKEN--GH---KVTDTEYIPY 273

Qy    299 VTLMKQVPGAGQOPAIOAEYSYTS-HNVVGGGSNGI-WNNKLDNLYGLMTEYNYGSTESRR 356
               |:  |  :|  ||  :   |  || ||| ||    |:||::  ||::| |:|:|:||
```

```
Db  1299  VTGWYFLGNGYRVFNPVLMRFHSPDSWSPFGRGGINPYTYCQGDPINRIDLNGHLSAGGI  1358
Qy  1390  TGIGMGIAGLLLTIAT--GMAIA----AAGGIAAAIASTTALAFGALSVTSDITSIV     1443
Db  1359  LGIVLGAIGIIVGIVSLGAGAAISAGLIAAGGALGAIASTSALAVTATVIGLAADSIGIA  1418
Qy  1444  SGALEDASPKRASSIIGWSMGMGAAGLAESAIKGGTKLATHLGAFAEDGENALLKST-SE  1502
Db  1419  SAALSEKDPKTSGIINWISAGLGVLSFGISAI----------TFT----SSLVKSARSG   1463
Qy  1503  SSRIKWGVTRSLDREIVRNEEGQVIKDHSR  1532
Db  1464  SQAVSAGVIGSVPLEF-----GEVASRSSR  1488
```

RESULT 2
US-09-492-709A-340
; Sequence 340, Application US/09492709A
; Patent No. 6720139
; GENERAL INFORMATION:
; APPLICANT: Zyskind, Judith
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Trawick, John
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Froelich, Jamie M.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; TITLE OF INVENTION: ESCHERICHIA COLI
; FILE REFERENCE: ELITRA.001A
; CURRENT APPLICATION NUMBER: US/09/492,709A
; CURRENT FILING DATE: 2000-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 340
; LENGTH: 1426
; TYPE: PRT
; ORGANISM: E. Coli
US-09-492-709A-340

Query Match   3.8%; Score 339; DB 4; Length 1426;
Best Local Similarity   20.2%;
Matches 257; Conservative 150; Mismatches 409; Indels 458; Gaps 60;
Pred. No. 1.1e-16;

Qy   572 KFTYTKQDDTLQQSHSITTHDNFTIHRSQVRSRYTGRLFSDTDTKDIVTQMSYDKLGRLL 631
Db   317 RYTYTEAGELL----AVYDRSNTQVRAFTYDAQHPGRMVAHRYAGRPEMRYRYDDTGRVV 372
Qy   632 TRTLNSGTPYANTLTYDYELNNLQDDNRPPFVITTTD----------------------- 668
Db   373 EQLNPAG--------LSYRY-----LYEQDR----IIVTDSLNRREVLHTEGAGLKRVVKKE 418
Qy   669 -VNGNQLRNEFDGAGRHVSQCLKDSDGDGKFYIIHTQQYD-----EQGRHHTSTYSDYL 721
Db   419 LADGSVTRSGYDAAGRLTAQ--TDAAGRRTEYGLNVVSGDITDITTPDGRETKFYND--- 474
Qy   722 TNGRQQT----DPDKVHLSMSKSYDNWGQIANTHWSGYGVSEKITVDPIITLTATKQLQSNSN 778
Db   475 --GNQLTAVVSPD--GLESRREYDEPGRLV-------------------------SE 502
Qy   779 NVQTGKEVT-TYTPSQQPIQITLFDEAGHLQSCHTLTRDGWDRVRKETDAIGQCTIYQYD 837
Db   503 TSRSGETVRYRYDDAHSELPATTTDATG--STRQMTWSRYGQLLAFTDCSGYQTRYEYD   559
Qy   838 NYNRVIQITLPDGTIVNRKYAPFSTDTLITDIRVNGISLGQQTFDGLSRLTQSQDG-GRV 896
Db   560 TAVHRE--------------------------EGISL-YRRYDNRGRLTSVKDAQGRE   595
Qy   897 WAYTYSAGNDQCPSTVITPDGQFIHYQYQ-------------------------- 925
Db   596 TRYEYNAAGDL--TAVITPDGNRSETQYDAWGKAVSTTQGGLITRSMEYDAAGRVISLTNE 653
```

```
Db   274  AAALTIQPGNGQPAVSKSYEYSSVVHNFLGYSSSGRTSFDSSQDNLYLVTGKYTYSSIE---   330
Qy   357  YKDKBGHDQIVRIERTYNNHLLTSECKQONGYIQTTETAYYALIGHNFDSOPSFQOLP-     415
Db   331  -RVLDGQSVVSVIERVFNKFHLMTKEAKTQDNKRITTEITYNEDLSKSFSFSEQPENLQQPS   389
Qy   416  KTKIETWRSADNSYRSEITETTFDESGNPLTKVIKDKKTQKIISPSTHWEYYPPAGEVDN    475
Db   390  RVLTRYTDIQTNTSREETVNIKSDDWGNTLL-ITETSGIQX------EVYYYPVNGEGNS    442
Qy   476  CPPEPYGFTRFVKKIIQTPYDSEFKDDPEKFIQYRY-----SLIGSQ-----SHVTLKIEE  526
Db   443  CPADPLGFSRFLKSVTQKGSPDAAQSVANKVIHYTYQKPFTTGAYVKEYVSKVSETIDN    502
Qy   527  RHYSATQLLNSTLFQYNTDKSELGRLLKQTECTKGENGKTYSVVHKFTYTKQDDTLQQSH   586
Db   503  KIARTFSYVNSP-----TSKSH-GSLAKITSVMMNQ-----QTVTTFKYEYSSSEMTTNA  551
Qy   587  SITTHDNFTIHRSQVRSRYTGRLFSDTDTKDIVTQMSYDKLGRLLTRTLNSGTPYANTLT  646
Db   552  TVTGFDGAHMESKNVTSIYTHRQLRKVDVNHVITDQSYDLLGRITGQIIDPGTAREIKRN   611
Qy   647  YDYELNNLQDDNRPPFVITTTDVNGNQLRNEFDGAGRHVSQCLKDSDG-------DGKF    698
Db   612  YVVQYPGGDENDFWP-VMIEVDSQGVRRKTHYDGMGRICSIERQDDDGAWGTSGIYQGTY  670
Qy   699  YTIHTQQYDEQGRHHTSTYSDYLTNGRQQTDP-----DKVHLSMSKSYDNWGQIANTHWS  753
Db   671  RKVLARQYDVLGQLSKEISNDWLWN--LSANPLVRLAITPLVTTKKTYKYDGWGNLYSTEYS  728
Qy   754  YGVSEKITVDPIITLTATK------QLQSNSNNVQTGKEVTTYTPSQQPIQITLFDEAGHL  807
Db   729  DGRIELEIHDPITRTITQGVKGLGMLNIQQNNF----------EQPASIKAVYPDGTI    776
Qy   808  QSCHTLTRDGWDRVRKETDAIGQCTIYQDYNRVIQTLPDGTIVNRKYAPFSTDTLIT     867
Db   777  YSTRTYRYDGFGRIVTETDAEGHATQIGYDVFDRIVKKTLPDGTILESSAYASFSHEELIS  836
Qy   868  DIRVNGISLGQQTFPDGLSRLTQSQDGGRVWAYTYSAGNDQCPSTVITPDGQFIHYQYQPE  927
Db   837  ALNVNGTQLGALVVDGLGRVISDTDVGGRKTEYLYGPQGDK-PIQSITPSHNKQMDYLIYY  895
Qy   928  LDDAVLQVASNEITQQFSYNPVTGALLKAVAEGQSLTPI-YYPSGRLKMEN-INDMKKMS   985
Db   896  LGSVMSKFTTGTDQQNFRTHSKTGTLLSA-SEGVSQTNVSYFPSGVLQRESFLRDNKPIS   954
Qy   986  ---YLWTLRGLENGYTDLTGTIQKISRDTHGRVTQIKDSSIKTTLNYDDLNRHIGSQVTD  1042
Db   955  SGEYLYTMSGLIQRHKDSFGHNHVSYDAQGRLVKTEQDAQYATFEYDNVGRLITTTTKD   1014
Qy  1043  LATGHMLITTVEFDGLNREIGRKLCDSSGHTLDIQQSWIKTQQLANRIVKLNGVLQRTEQ  1102
Db  1015  TTSLSQLVTKIEYDAPDREIKRSLISDFSIQV-ITLSYTKNNQISQRITSIDGVVMKNER  1073
Qy  1103  YSVDSRNRLNQYKCDGAECPTDKYGHSIVTQNFTYDIYGNITACHTTFADGTEDHATFKF  1162
Db  1074  YQYDNNQRLRSQYCEGEQOSPIDHTGRVLNQQIYHYDQWGNIKLKDNTYRDGKET-VDYHF  1132
Qy  1163  ANPTDPCQLTEVHHTHPDMPDNIRLKYDKAGRVINIT-DNHGNTENFTYDTLGRL---QN  1218
Db  1133  SQ-ADPTQLIRI----TSDKQQIELSYDANG---NLTRDEKGQT--LLYDQNNRLVQVKD  1182
Qy  1219  GQGSV---YGYDPLNRLVSQ--KTDTLDCELYYRETMLVNEVRNGEMIRLLRTGETIIAQ  1273
Db  1183  RLGNLVCSYQYDALNKLTAQVLANGTVNRQ-HYASGKVTNIQLGDEAITWLSSDKQRIGH  1241
Qy  1274  QRA----SKVLLTGTDSQQSVLLTSDKQNLSQEAYSAYGKHKSTANDASILGYNGERADP  1329
Db  1242  QSAKNGQSYYQYQYGIDHNSTVIASQNENELMALSYTPYGFRSLI---SSLPGLNGAQVDP  1298
Qy  1330  VSGVTHLGNGYRSYDPTLMRFHTPDSLSPFGAGGINPYSYCLGDPINRSDPSCHLSWQAW  1389
```

Query Match   3.7%; Score 331.5; DB 4; Length 1377;
Best Local Similarity 20.8%; Pred. No. 4e-16;
Matches 288; Conservative 170; Mismatches 450; Indels 477; Gaps 75;

```
Qy   211 LFPGQKEGY-RTELRFLNRQLNSIHNFSLGNENPLTWSFGYTPIGKNGILGQWITSMTAP 269
             ||||: |     :|: :   |  :   : :::  ::
Db   126 LFPGE-DGYSRSESLMIVR--GGVAKLDEGHRLAALWQ--------------------- 160

Qy   270 GGLKETVNYS---------NNNQG------PVLPYVTLMKQVPGAGQ- 310
              :::              :  ::|    |
Db   161 -ALPEELRLSPHRYLATNSPQGPWWLLGWCERVPEADEVLPAPLPPYVRVLTGLVDRFGRT 219

Qy   311 PAIQAEYSYTSHNYVGGGSNGIWNNKLDNLYGLMTEYNYGSTESSRYKDKEGHDQIVRIE 370
                                              :        :::
Db   220 QTFHREAAGEFSGEITGVTDGAWRH------FRLVL--TTQAQR- 255

Qy   371 RTYNNYHLLTSECKQQNGYIQTTETAYYAIIGHNFDSQPSQFQLPKT---KTETWRSADN 427
             :                                           ::
Db   256 ----------AISG--GTEPSAF--PDTLPGYTEYGR--DN 287

Qy   428 SYRSEITETTFDESGNPLTKVIKDKKTQKIISPSTHWEYPPAGEVDNCPPEP---YGFT 484
                                                              ::
Db   288 GIRLSAVWLTHDPB------------------YP----ENLPAAPLVRYGWT 317

Qy   485 --------RPVKKIIQTPYDSEFK----------DDPEKFIQYRYSLIGSQSHVTLKIE 525
                                                  :
Db   318 PRGELAVVYDRSGKQVRSFTYDDKYRGRMVAHRTGRPE--IRYRYD--SDGRVTEQLN 372

Qy   526 ERHYSATQLLNSTLFQYNTDKSELGRLLKQTEC--TKGENGKTYSVVHKFTYTKQDDTLQ 583
                                                  :::
Db   373 PAGLSYT----------YQYEKDRITITDSLDRREVLHTQGEAG-LKRVVKK- 413

Qy   584 QSHSITTHDNFTIHRSQVRSRYTGRLFSDTDTKDIVTQMSYDKLGRLLTRTLNSGTPYAN 643
                                                  :::: :
Db   414 ------EHADGSVTQSQFDA--VGRLRAQTDAAGRTTEYSPDVVTGLITRIT--TPDGR 462

Qy   644 TLTYDYELNNLQDDNRPPFVITTTDVNGNQLRNEFDGAGRHVSQCLKDSD-----GDGKF 698
                                                  ::::
Db   463 ASAFYYNHHNQ------------LTSATGPDGLELRREYDELGRLIQETAPDGDITRYRYDNPH 514

Qy   699 YTIHTQQYDEQGRHHTSTYSDYLTNGRQQTDPDKVHLSMSKSYDNWGQIANTHWSYGVSE 758
                                                  :::
Db   515 SDLPCATEDATGSRKTMTWSRY---GQLLSFTDCSGVTRYDHDRFGQMTAVHREEGLSQ 571

Qy   759 KITVDPI-TLTATKQLQSNSNNVQTGKEVTTYTPSQOPIQITLFDBRAGHLQ------SC 810
                                                  ::
Db   572 YRAYDSRGQLIAVKDTQGHETRYE-----------YNIAGDLTAVIAPDGSR 612

Qy   811 HTLTRDGWDRVRKETDAIGQCTIYQYDYNYRVIQITLPDGTIVNRKYAPFSTDTLITDIR 870
                                                  :::
Db   613 NGTQYDAWGKAVRTTQG--GLTRSMEYDAAGRVIRLTSENGS---HTTFRYDVLDRLLIQ 666

Qy   871 VNGISLGQQTF--DGLSRLTQSQDGGRVWAYTYSAGNDQCPSTVIITPDGQFIHYQYQPEL 928
             :  ||::                                    :
Db   667 ETGFDGRTQRYHHDLTGKLIRSEDEGLV-------------THWHYD-EA 702

Qy   929 DDAVLQVASNEITQQFSYNPVTGAL--LKAVAEGQSLTPIYY---PSGRL--KMENINDM 981
             :::  :||: :  ::  |:|:             | |:    : :
Db   703 DRLTHRTVKGETAERWQYDE-RGWLTDISHISEGHRVA-VHYRYDEKGRLTGERQTVHHP 760

Qy   982 KKMSYLW------TLRGLENGYTDLTGTIQKISRDTHGRVTQIKDSSIKTLNYDDLN 1033
                                              :::
Db   761 QTEALLWQHETRHAYNAQGLAN----RCIPDSLPAVEWLTYG-------SGYLAGMKLGD-- 809

Qy   926 --------PELDDAVLQVASNEITQQFSYNPVTGALLKAVAEGQSLTPIYYPSGRLKME 976
                :::  :||: : ::    :|  |:
Db   654 NGSHSVFSYDALDRLVQQGGFDGRTQRYHYD-LTGKLTQSDEGLVILWYDESDRITHR 712

Qy   977 NINDMKKMSYLWTLRGLENGYTDLTGTIQKISRDTHGRVTQIKDSS----IKTTLNYDDL 1032
             ::  ||:                                       ::
Db   713 TVNGEPAEQWQY----------DGHGWLTDISHLSEGHRVAVHYGYDDK 751

Qy  1033 NRHIG--SQVTDLATGHML------- 1049
             :::  ::   |: :  :
Db   752 GRLTGBECQTVENPETGELLWQHETKCHAYNEQGLANRVTPDSLPPVEWLTYGSGYLAGMKL 811

Qy  1050 --TTTVEF--DGLNREIGRKLCDSSGH--TLDIQQSWLKTQQLANR----------- 1089
             ::|        :|:
Db   812 GGTPLVEYTRDRLHRRTVRSFGSMAGSNAAYELTSYTPAGQLQSQHLNSLVYDRDYGWS 871

Qy  1090 --IVKLNGVLQRTEQYSYDSRNRLNQYKCDGAEC------PTDKYGHSI---------- 1130
                                                  :::
Db   872 DNGDLVRISGPRQ-TREYGYSATGRLESVRTLAPDLDIRIPYATDPAGNRLPDPELHPDS 930

Qy  1131 -----VTQN---FTYDIYGNITACHTTFADG--TEDHATFKFANPTDPCQLTEV 1174
             :::  :   ::
Db   931 TLTWPDNRIAEDAHVYRHDEYGRLTEKTDRIPAGVIRITDDERT------------ 975

Qy  1175 HHTHPDMPDNIRLKYDKAGRVINITD-NHGN--TENFTYDTLGR------ 1215
             ::::    :::|  :
Db   976 HHYH---------YDSQHRLVFYTRIQHGEPLVESRYLYDPLGRRMAKRVWRRERDLTG 1025

Qy  1216 -LQNGQGSVYGVGYDPLNRLVSQKTDTLDCELYR------ 1247
             ::|
Db  1026 WMSLSRKPEVTWYGWDG-DRLTTVQDTTRIQTVVEPGSFTPLIRVETENGEREKAQRRS 1084

Qy  1248 --ETMLVNEVRNG--------EMIRLL-RTGETIIA---------QQRASKV 1279
             ::                                   ::|
Db  1085 LAETLQQEGSENGHGVVFPAELVRLLDRLEEEIRADRVSSESRAWLAQCGLTVEQLARQV 1144

Qy  1280 LLTGTDSQQS-----------VILTSDKQNLSQEA-YSAYGKHKSTANDASILGYN--- 1323
             :::
Db  1145 EPEYTPARKAHLYHCDHRGL-PLALISEDGNTAWSAEYDEWNGQLNEENPHHV--YQPYRL 1202

Qy  1324 -GERADPVSGVTHLGNGYRSYDPTLMRFHTPDSLSPPG-AGGINPYSYCLGDPINRSDPS 1381
             :::
Db  1203 PGQQHDEESSGLYY--NRHRYDPLQGRYITQD---PMGLKGGWNLYQYPL-NPLQQIDPM 1256

Qy  1382 GHLSWQAWTGIGMGIAGLLLTIATCGMAIAAAGGIAAAIASTSTTALAFGALSVTSDITS 1441
             :::
Db  1257 GLL--QTWDDARSG-----ACTCGV----CGViSRIIGPSKFDSTADAALD------- 1296

Qy  1442 IVSGALEDASPKASSILGWVSMGMGAAGLAESAIKGGTKLATHLGAFAEDGENALLKSTS 1501
             :::
Db  1297 ----ALKETQNRS-------LCNDMEYSGIVCKDTNG---------KYFASKAETDNLR--K 1336

Qy  1502 ESSRIKWGVTRSLDREIVRNEEGQVIKDHSRGYTDNFMGKGEQAILVHGDK--DGFLYHT 1559
             :::
Db  1337 ESYPLKRKCPTGTDRVAAYHTHG---ADSHGDYVDEFFSSSDKNLVRSKDNNLEAFYLAT 1393

Qy  1560 EGNKH--NGKGPY 1570
             ||
Db  1394 PDGRFEALNNKGEY 1407
```

RESULT 3
US-09-711-164-467
; Sequence 467, Application US/09711164
; Patent No. 6589738
GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERET
; FILE REFERENCE: ELITRA.008A
; CURRENT APPLICATION NUMBER: US/09/711,164
; CURRENT FILING DATE: 2000-11-09

```
Qy   1034 RHIGSQVTDLATGHMLTTTVEF--DGLNREIGRKLCDSSGHTLDIQQSWLKTQQLANRIV 1091
          ||| ::|| :   ||||          :   ||   |   |  :::
Db    810 -------------TPLVEYTRDRLHRETLR-----SFGRYELTTAVTPAGQLQSQ-- 846

Qy   1092 KLNGVLQRTEQYSYDSRNRLNQYKCDGAECPTDKYGHSIVTQNFTYDIYGNITACHTTFA 1151
          |:::|          :   |||||            |:::::
Db    847 HLNSLLS-DRDYTWNDNGELIR-----ISSPRQ-------TRSYSYSTTGVHTTAA 893

Qy   1152 DGTEBDHATFKFANPTDPC--QLTEVHHTHPD-------IRLKYDKAGRVIN 1197
          ||                   |||          :::  ||| |
Db    894 -----NLDIRIPYATDPAGNRLPD-PELHPDSTLSWPDNRIARDAHYLRYDRHGRLTE 947

Qy   1198 ITD--------NFGNTENFTYDTLGRLQNGQ---GSVYGYDPLNRLVSQKTDT 1239
          :||           ::    ||   |      ::||||:
Db    948 KTDLIPEGVIRTDDERTHRYHYDSQHRLVHYTRTQYEEPLVESRYLYDPLGRRVAKRVWR 1007

Qy   1240 LDCEL------YY-----RETMLVNE------VRNGEMIRLLR----TGETI 1270
          :                 :   :           :  |
Db   1008 RERDLTGWMSLSRKPQVTWYGWDGDRLTTIQNDRTRIQTIYQPGSFTPLIRVETATGELA 1067

Qy   1271 IAQQRA--------------------SKVLLTGTDSQQSVIL-------TS 1294
          |:||                        :  |         :|:||
Db   1068 KTQRRSLADALQQSGGEDGGSVVFPPVLVQMLDRLESBILADRVSEESRRWLASCGLIVE 1127

Qy   1295 DKQN--------LSQEA----YSAYGKHKSTAND---AS 1318
          |||         :||:        ::||| |
Db   1128 QMQNQMDPVYTPARKIHLYHCDHRGLPLALISKEGTTEWCAEYDEWGNLLNEENHQLQQ 1187

Qy   1319 ILGYNGBRADPVSGVTHLGNGYRSYDPTLMRFHTPDSLSPFG-AGGINPYSYCLGDPINR 1377
          ||||          :   :: ||||                 ||||
Db   1188 LIRLPLGQQYDEESGLYY--NRHRYYDPLQGRYITQD-----PIGLKGGWNFYQYPL-NPVTN 1241

Qy   1378 SDPSG 1382
          :|| |
Db   1242 TDPLG 1246
```

```
RESULT 4
US-09-252-991A-28424
; Sequence 28424, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT:  Marc J. Rubenfield et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  107196.136
;  CURRENT APPLICATION NUMBER:  US/09/252,991A
;  CURRENT FILING DATE:  1999-02-18
;  PRIOR APPLICATION NUMBER:  US 60/074,788
;  PRIOR FILING DATE:  1998-02-18
;  PRIOR APPLICATION NUMBER:  US 60/094,190
;  PRIOR FILING DATE:  1998-07-27
;  NUMBER OF SEQ ID NOS:  33142
; SEQ ID NO 28424
;   LENGTH:  998
;   TYPE:  PRT
;   ORGANISM:  Pseudomonas aeruginosa
US-09-252-991A-28424
```

```
  Query Match             3.6%;  Score  317;  DB 4;  Length 998;
  Best Local Similarity   22.0%;  Pred. No. 3.1e-15;
  Matches  222;  Conservative  129;  Mismatches  362;  Indels  296;  Gaps  47;

Qy    504 EKFIQYRYSLIGSQSHVTLKIEERHYSATQLLNSTLFQYNTDKSELGRLLKQTECTKGEN 563
          :                |   |::         :
Db    112 ERAWSYSYNALG-----LIERADGPRTDVQDVTLYAYD---------SRGNL 149

Qy    564 GKTYSVHKFTYTKQDDTLQQSHSITTHDNFTIHRSQVRSRYTG--RLFSDTDTKDIVTQ 621
          |||| :      :          ||| :: |||     :     |:
Db    150 TQVTNALGQVTRLGDYDERGKPGSITDANGVTSSLA---YTGVDGWLASVSTAGSTTR 204

Qy    622 MSYDKLGRLLTRTLNSGTPYANTLTYDYELNNLQDDNRPPFVITTDVNGNQLRNEFDGA 681
```

```
Qy    205 FDYDAVGQITRVTRGDG----SWLSYEY------DDARR--LVAIGNNLGERLEYDVDTK 252.
          || :::||  ||| |          ::||:      ||||
Db    682 GRHVSQCLKDSDGDGKFYTIHTQQ--YDEQR------HHTSTYS-DYILTNGRQQTDP 730

Qy    253 GNRTAQRIKDASGS----LVRQQOWAYDELGRLLRAVGAGGQTRSFAYDLNDNPVGETNP 308
          |||  |   |::          |||  |:|||  :         ::
Db    731 DKVHLsMSKSYDNWGQIANTHWSYGVSEKITVDPITLTATKQLQSNSNNVQTGKEVTTYT 790

Qy    309 RQ--FAHSQAFDALDRLVGQSDPLGGKTRLAYD------AQDNLTEVDPRGVTTRY 357
          |            |::|   ||  | |||         |
Db    791 PSQQPIQITLFDEAGHL-------QSCHTLTRDGWDRVRKETDAIGQCTIYQYDNYNRVI 843

Qy    358 E-------YDGLGNLTRLVSPDSGTTTFEHDAAGNVIRRTDARGAVTEYRYDALNRLV 408
          :           | ::|  :   :    :| |:  |          |||
Db    844 QITLPDGTIVNRKYAPFSTDTLITDIRVNGISLGQOTFDGLSRLTQSQDGGRVWAYTYSA 903

Qy    409 ERRSP---------SDPSLDVQYRYDLTADGNQGIGRLGAIBGARDS---TGETI 446
          :                   ||:|  ::|    ::| |  ||
Db    904 GNDQCPSTVITPDGQFIHYQXQPELDDAVLQVASNEITQQFSYNPVTGALLKAVAEGQSL 963

Qy    447 ------LVRYD-ERGNLVEQVRSIRLDQQTLLDRVTYRYDAA----NQL 485
                |:||    |||  |  :  |         :   ||
Db    964 TPIYYPSGRLKMENINDMKKMSYLWTLRGLENGYTDLTGTIQKISRDTHGRVTQIK----- 1019

Qy    486 LEIGYPS--------------GLAIGY---------PRNAGGQVASVTLAVG 514
          |   ||                ||  |             | |
Db   1020 DSSIKTT-------LNYDDLNRHIGSQVTDLATGHMLTTTVEFDG-----LNREIGRKLCDS 1069

Qy    515 DKAPSTLVGQIAYLPFGPLQR--------LTWGNGITLSREYDQDYQLLRQKVG- 560
          |      |  |                   | |      |   :|
Db   1070 SGHTLDIQQSWLKTQQLANRIVKLNGVLQR-------TEQYSYDSRNRLNQYK-CDGAEC 1121

Qy    561 --PWQSDYQ------HDANGNIQQHRHSLWGTLDYQYDPLDRLTEERGVQGG- 604
          |             |   :      :  | |         |   | | |
Db   1122 PTDKYGHSIVTQNFTYDIYGNIT-ACHTTPADGTEDHATFKFANPTDPCQLTEVHHTHPD 1180

Qy    605 ----RSYAYDAVGNRTQRSDNPASGGTASSQDYQYA 636
          :
Db                                                                  

Qy   1181 MPDNIRL--------KYDKAGRVINITDNHGNTENFTYDTLGRLQN-----GQGSVYGYD 1227
          ||:||               |  :
Db    637 -PDSNRLTAIGAQAVTSDAAG---NLTQDRA-ARKLAYDAQGRLQSVSLDGQQVAEYRYN 691

Qy   1228 PL-NRLVSQKTDTLDCELYRETMLVNEVRNGEMIRLLRTG-------ETTIAQQRA 1276
          |  :|           |              : ||::
Db    692 ALGQRIVKLTPESITTYLYGPDGQLLGEAEHDGSGRKLRAQYYLWLDSLPLATIDADYDA 751

Qy   1277 S--------KVLLTGTDSQQSVILTSDK--QNLSQEAYSAYGKHKSTANDAS--ILGYNGER 1326
          |         |||       |    |      | :  |:       |    ||||
Db    752 QGKVGNPTLYLHGDHLDTPRLATDASGQIAWQWQSDAFGREALSQGSTQVNLRFPGQY 811

Qy   1327 ADPVSGVTHLGNGYRSYDPTLMRFHTPDSLSPFG-AGGINPYSYCLGDPINRSDPSG--- 1382
          ||| ||||| |||| ||||| ||||  | : |   ||||||||||| ||| || |
Db    812 YDAESGLHY--NYFRDYDPETGRYVESD---PIGLSGGVNTYGYVQGAPLNRIDPLGLAA 866

Qy   1383 ---HLSWQAWTGIGMGI----AGLLLTIATGGMAIAAAGGIAAAIASTST 1425
          |       |  |:|       |||| |:| ||||  | | |
Db    867 IEIDIPKSAYDWIPGNIRLPAGRLL----GGVLLVASISGATPQADSDT 911
```

```
RESULT 5
US-09-328-352-6604
; Sequence 6604, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
;  APPLICANT:  Gary L. Breton et al.
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
;  TITLE OF INVENTION:  BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  GTC99-03PA
;  CURRENT APPLICATION NUMBER:  US/09/328,352
;  CURRENT FILING DATE:  1999-06-04
;  NUMBER OF SEQ ID NOS:  8252
; SEQ ID NO 6604
```

; LENGTH: 1627
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6604

Query Match   3.5%;   Score 313; DB 4; Length 1627;
Best Local Similarity   20.4%;   Pred. No. 1.5e-14;
Matches 303; Conservative 177; Mismatches 506; Indels 498; Gaps 73;

```
Qy    75  IGFGIGFNFGLSVYDRKNSLLSLSTGENY--KVIETDKTV-------------KLQ  115
Db   369  ITFAMGTEFFTHVDAQLGGIIQDSISRTYVSNLYQMDDAIFGARWVTPFTTKISRKFKYT  428

Qy   116  QKKLDNLRFEKDLKENCYRIIHKSGDIEVLTGFNNNAPDLKVPKKLLNPAGHAIY---ID  172
Db   429  SKKKDHKDYLNGLEYIC-------LDGRAIDLPDLKK-----GQSIYDPIEQ  468

Qy   173  WNFEATQPRLNRI----------YDDLDGHDIPLLNLEYQGLIKTILTLFPCQKEGVRTELR  224
Db   469  YTYTVLSDQLHLIAYGEDEKRYYEKYGEDYRLSYIB--------------RKNGFKVALR  514

Qy   225  F-----LNRQLNSIHNFSLGNENPLTWSFGVTPIG---------KNGILGQWITSMTAPGG  271
Db   515  YDHVSTDNKTILSDILFKQDDNLLAHLALQLTPQGLVSDIWTIKNGQLDRVLAS-------  568

Qy   272  LKETVNYSNNNQGHHFPQSANLPVLPYVTLMKQVPGAGQPAIQAEYSYTSH--------NYV  325
Db   569  ------YDYDQQG-------------DLVQATNEFAASYYQYTHHLLITRYTDLT  604

Qy   326  GGGSNGIWNNKLDNLYGLMTEYNYGSTESRRYKDKEGHDQIVRIERTYNNYHLLTSECKQ  385
Db   605  HRGMNLKWDGILPTSKAIEWADNASRASKLEWDK----NIRKT----TVLDVEGNS  653

Qy   386  QNGYIQTTETAYYAIIGHNFDSQPSQFQLPKTKTETWRSADNSYRSEITETTFDESGNPL  445
Db   654  TEHYDIDGYTYRIVYPDNFE---BCFFRRDAKNITLHIAKDGSK---TSYTYDERGNVL  707

Qy   446  TKVIKDKKTQKIISPSTHWEYYPP---AGEVDNCPPEPY---GFTRFVKKIIQTPYDSE  498
Db   708  TTTQDDGAT------SYFEYDEKNQLTGMVDAEQGRWFKQYDGSGNLIKEI--------  752

Qy   499  FKDDPEKFIQYRYSLIGSQSHVTLKIEERHYSATQLLNSTLFQYNTDKS---ELGRLLK  554
Db   753  --DPLK----HETAYV--YNAMGLVTSITDAKGGSKSSLKYDDQGNLIS  792

Qy   555  QTEC----TK---GENGKTYSVVH----KFTYTKQDDTLQQ--------  584
Db   793  YTDCSGKETKWQYDERGRVISIENALNQKVEYFYTELTLENREPIIKGLPLNAFGQLEKI  852

Qy   585  SHSITTHDNFTIHRSQVRSRYTGRLFSDTDTKDIVTQMSYDKLGRLLIRT--LNSGTPYA  642
Db   853  KHADGTEEHF-IHDAE-------GRILAHVDPKQNITRYEYDEAGLILSRTDALN-----  899

Qy   643  NTLTYDYE----LNNLQDDNRPPFVITTTDVNGNOLRNEFDGAGRHVSQCLKDSDGDGKF  698
Db   900  HKLKYKWDRLGRLTRLINE---------NGASYQFFYDVASRLV---KEIDFDGKE  943

Qy   699  YTIHTQQYDEQ-GRHHTS--TYSDYLTNGRQQTDP-DKVHLSMSKSYDNWGQIANTHWSY  754
Db   944  TVYH---YDEKSGQLATSIEVASAYGQDLKDRAAPKDRIQQFI---FDSMGRLEQRTAGY  997

Qy   755  G------VSEKITVDPI-TLTATKQLQSNSNNVQTGKEVTYYTPSQQPIQITLFDEA  804
Db   998  GHYGLELEEKQTEFFAYDVMGRIIQAKNAQSNLQ-----------WFYDAA  1037

Qy   805  GHLQSCHTLTRDGWDRVRKETDAIGQCTI--YQYDNYNRVIQITLPDGTIVNRKYAPFST  862
Db  1038  GNLVQEH----QDDYKINKTAVWKHQYDEINDRIKTTRPDGQVID-------------1078

Qy   863  DTLITDIRVNGISLGQQTFDGLSRLTQSQDGGRVWAYTYSAGNDQCPSTVITPDGQFIHY  922
Db  1079  --------W-LTYGSGHVQ---SLIVNGQDFVSF  1100
```

```
Qy   923  QYQPELDDAVLQVASNEITQQFSYNPVTGALLKAVAEGQSLTPIYPSGRLKMENINDMK  982
Db  1101  ----BRDDLHREIA--------------RHYANGVSQEQQYDLAGRLKSQMM----  1134

Qy   983  KMSYLWTLRGLENGYTD---------------LTGTIQKISR---DTHGRVTQIKDSSI-KTTLNY  1029
Db  1135  ------LSEHENGYQNQYKRHNNALEQTSQLVQRLYQYDKTGELTAIRDTRRGNIAYKY  1187

Qy  1030  DDLNRHIGSQVTDLATGHMLTTVEFDGLNRREIGRKLCDSSGHTLDIQQSWLKTQQLA--  1087
Db  1188  DPVGRLLE---------ASSKLGKETFSFDPASNIL---DSYHSQKVQSHSQKLDETSYG  1235

Qy  1088  -NRIVKLNGVLQR--TEQYSYDSRNRLNQYKCDGAECPTDKYGHSIVTQNFTYDIYGNIT  1144
Db  1236  YNRLV--NNVVKEYLDQQYQYDAYGQLIRQKTSQGDL---------NLEWDVYGRMV  1281

Qy  1145  ACHTTFADGTEDHATFKFANPTDPCQLTEVHHTHPDMPDNIRLKYDKAGRVINITDNHGN  1204
Db  1282  KSRNS--------------QYTAEY---------RYDALGRRIQKWSKHHH  1309

Qy  1205  TENFTYDTLGRLQNQGQSVYGYDPLNRLVSQKTDTLDCBLYRETMLVNEVRNGEM--IR  1262
Db  1310  T-------GQEQN---IIYGWDG-DTLAYESTEEJLTKHYIYEKDSFVPMLQAVYLSPIE  1357

Qy  1263  LLRTGETI------IAQQRASKVLLTGTDSQOSVILTSDKQNLSQE----------AYS  1305
Db  1358  LHQTPDWSDRPYNIHRDPLWKTEKEGKEFPDDWYHCDHLGTPQEMTDHTGAIIWKAEYK  1417

Qy  1306  AYGKHKSTANDASI----------LGYNGERADPVSGVTHILGNGYRSYDPTLMRFHTPDS  1355
Db  1418  AWGECKAEEAKSNFPENSEIISNNIRFQGQYFDEETGLHY--NRYRYYSPYVGRFVSKD-  1474

Qy  1356  LSPFG-AGGINPYSYCLGDPINRSDPSGHLSWQAWTGIGMGIAG  1398
Db  1475  --PIGLLGGNNVVYTA-KNPITWIDSKGLCSTTLNRNLG-GVKG  1514
```

RESULT 6
US-09-543-681A-5329
; Sequence 5329, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABI
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 5329
; LENGTH: 1586
; TYPE: PRT
; ORGANISM: Proteus mirabilis
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (1576)
; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
US-09-543-681A-5329

Query Match   3.4%;   Score 306; DB 4; Length 1586;
Best Local Similarity   22.1%;   Pred. No. 4.9e-14;
Matches 243; Conservative 129; Mismatches 385; Indels 342; Gaps 59;

```
Qy   499  FKDDPEKF----------IQYRYS----LIGSQSHVTLKIEERHY-SATQLLNS-  537
Db   486  FEANPHKINTLRLMKSADRHNNALHYRYANDGELVQIHDDAYLTDIRLHYDEITQRLQSV  545

Qy   538  --------------TLFQYNTDKSELGRLLKQTECTKGENGKTYSVVHKFTYTKQDDTL---QQS  585
Db   546  TRHQGQEEKILVTTYDAQQ--RLVQVTNADK---------RVTRRFGWDDESGLMAMHQYA  596
```

GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABIL
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 7560
; LENGTH: 1439
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-7560

```
Query Match             3.3%;    Score 294;  DB 4;   Length 1439;
Best Local Similarity  20.3%;    Pred. No. 3.6e-13;
Matches  251;   Conservative 141;  Mismatches 386;   Indels 460;  Gaps 62;

Qy   347  YNYGSTESRRYK-------DKEGHDQIVRIERTYNNYHL------LITSECKQONGYIQ  391
          :||| :  || |       |:                 |     |::   |   ||:
Db   376  YHYPPFDDKGYSRLSHIVDNVGNEQRFH----YNEHHQLIHITGCGDLNIECEYQS--FQ  429

Qy   392  TTE------TAYYAIIGHNFDSQPSQF---------QLPKTKTET--------------WR  423
          |||            || |:|| :            :|  |:|:              ||
Db   430  LAEKTVSRLTAVYQVNPHQIRRRLCAYFYNESAQLIRVEQQTNHPYRQFGWTDAGVMAWH  489

Qy   424  SADNSYRSEITETTFDEGSNPLTKVIKDKKTQKILISPSTHWEYYPPAGEVDNCPPEPYGF  483
                                               :  ::     GESYRLEYDDINL  531
Db   490  SDKYGLRSEYRWALSEDN----LWRVIENKTSE-------

Qy   484  TRFVKKIIQTPYDSEFKDDPEKFIQYRYSLIGSQSHVTLKIEERHYSATQLL--------  535
          |                                    |::   |  |  |
Db   532  TR------TAY--WHDGSTSFWQLNH-----------DHQIIHYIDRTGIKTALIWDEFGLPC  575

Qy   536  ---NSTLFQYNTDKSELGRLLKQTECTKGENGKTYSVVHKFTYTKQDDTLQQSHSITTHD  592
             |                               :                    IT--
Db   576  GCRNALGHTHISEWDALGRLLSITD----GNGNQT--------RWQYQNERERL-------  619

Qy   593  NFTIHRSQVRSRYTGRLFSDTDTKDIVTQMSYDKLGRLLTRTLNSGTPYANTLTYDYELN  652
                                 VFWPDNTE---SRIAYDSIGRLIKEI------SPLHQITEYRYDFK  656
Db   620  ---------

Qy   653  NLQDDNRPPFVITTDVNGNQLRNEFDGAGRHVSQCLKDSDGDGKFYTIHTQQYDEQGRH  712
          |    | |             :   |  || |        |   ||   |  ::|| |
Db   657  TTL---RP----TIRIDAKQGRSEFLWNKR--GQLLRHTDCSGK---QHIWCYDDEGRV  703

Qy   713  HTSTYSDYLTNGRQQTDPDKVHLSMSKSYDNWGQIANTHWSYGVSEKITVDPITLTATKQ  772
          |||          ||                 :      :        ILPD----NSTVQ  736
Db   704  VSQ-------TNALQE---------ATEYQYDEVGHLARI---------

Qy   773  LQSNSNNVQTGKEVTTYTPSQQPIQITLFDEAGHLQSCHTLTRDGWDRVRKETDAIGQCT  832
          |         |                          |:               NAFGRVTEIDKLARHI  777
Db   737  LAWNNAAGLLTHHQRNDNTPCQWQY-------

Qy   833  IYQYDNYNRVIQITLPDG--TIVNRKYAPFSTDLITDIRVNGISLGQQTFDGLSRLTQS  890
          | |  |  |  ||  |:      | |   |  |    :  ||  || |::  |  |
Db   778  HYHYNAQGALISIENANGGRYLLNRD----AEDRLVEBIRPDE-TLLQYTYNVAGRLVEB  832

Qy   891  QD-GGRVWA----------YTYSAGNDQCPSTVITPDGQFIHYQYQPELDDAVLQVASNEITQ  942
          ||            |                        :|| | |  |
Db   833  AHLGDRVFTSAPRTILLDYDAAGNLVKRETLTD------RYQYQWDSMNRLL-VASKQPNQ  886

Qy   943  ---------------------QFSYNPVTGALLKAVAEGQSLTPIYYPS  970
Db   887  RGLEMGLQANQVHFTYDALGRIIREQTGDDIVEPNYDELNNLSRLTLPQGDSLNWLYYGS  946

Qy   971  GRLKMENINDMKKMSYLWTLRGLENGYTDLTGTIQKISRDTHGRVTQIKDSSIKTTLNYD 1030
          :|                        |||:
Db   947  GHATAIN--------HLVDSR------SQLITEFERD.969

Qy  1031  DLNRHIGSQVTDLATGHMLTTTVEFDGLNREIGRKLCDSSGHTLD---IQQSWLKTQQLA 1087
```

(Second column — continued alignments)

```
Qy   586  HSITTH------DNPTIHRSQVRSRYTGRLFSDTDTKDIVTQMSYDKLGRLLTRTLNSGT  639
          : | |       |||| |  || |      ||  ||  ||::   :
Db   597  TGVSSHYRWQRFDAFTIEDNEPEWRVVEHWLKDGKRCLEHTELITYDLAQRTLTTVETGE  656

Qy   640  PYANTLTYDYELNNLQDDNRPPFVITTTDVNGNQLRNEFDGAGRHVSQCLKDS---DGDG  696
                     :  |||||        TSRLLKKAIAPDGSE  699
Db   657  ---------TTFRRWNEQQQIEYTNALNETWWFEWD----

Qy   697  KFYTIHTQQYDEQGRHHTSTYSDYLTNGRQQTDPDKVHLSMSKSYDNWGQ------IANT-  750
          ||||  ||                 :||||||||||||||||||||||||
Db   700  WGYT------YDERG-----NLTQWTDPEQ---QSTCYD-WDKDFAPPTAQTL  737

Qy   751  ------HWSYGVSEKI--TVDP---ITLTATKQLQSNSNNVQTGKEVTTYTPSQQPIQIT  799
          ||                 :||        ||
Db   738  PNGAAWHWEYNEHGDIRRVIDPLGHITRLAWDDQGLCLGQVDAKGNETHYRYNARGQLIE  797

Qy   800  LFDEAGHLQSCHTLTRDGWDRVRKETDAIGQCTIYQYDNYNRVIQITLPDGTIVNRKYAP  859
          |    |   |:    |    |  :|              :  :    || |  :
Db   798  QRDCSGYPT---TLTYDDWGQLRSLTNAQNETTYTYFSEAGLLLTECLLPDGT--ENRYDY  852

Qy   860  FSTDTL--TTDIRVNGISL-----GO---------QTFDGLSRLTQSQDGRV  896
          |   |                  |                       :
Db   853  DATGQLVGITDAGERHILLRRNRRGQVIARRDPAGHWLHFHYDTFGRMQALENEQ--GEQ  910

Qy   897  WAYTYSAGNDQCPSTVITPD----GQFIHYQYQ--------PELDDAVLQVASNE  939
          |   ||  |               |  :|  |          |
Db   911  YRFEYDALHR-------LTDEHDLIGQQKHYQYDVWGNVTGIKTTPGPSIDTPI--PLSPQ  962

Qy   940  ITQQFSYNPVTGALLKAVAEGQSLTPIYYPSGRLKMENINDMKKMSYLW------TLRGLE  994
                                             GRILLFRENADY-RTEYLYQPLSVTLRRRVP  999
Db   963  VT--TFGYDKV------

Qy   995  NGY---TDLTGTIQKISRD-----THGRVTQ-IKDSSIKTTL--NYDDLNR---------- 1034
          ||:           |:        |          ||
Db  1000  MAVWHEAERTGTTARVEYQDALFFTYDKVGQLVREASARGDYQHHYDVLGNITRTELPHQ 1059

Qy  1035  -----HIGS---QVTDLATGHMLTTTVEF--DGLNREIGRKLCDSSGHTLDIQQSWLKT 1083
                 :|            |  |        |   |                :
Db  1060  RAFEYLYYGSGHLQQTQWRDNEQLTVLAEYQRDRLHRETLR----TSG-ALDNETGYDCR 1114

Qy  1084  QQLANRIVK--------LNGVLQRTEQYSYDSRNRL----------NQYK 1115
                        :|                                :::
Db  1115  GRITHQVARQMNASQFVTPVIDR--RYRWDKRNQLIERSVSYGQTGEVFTAGHWYYHSYQ 1172

Qy  1116  CDGAECPTPKYGHSIVTQNFTYDYGNITACHTTPADGTEDHATFXFANPTDPCQLTEVH 1175
          |              :                  :::
Db  1173  YDPIGQLTAHLG-SVQTEHIFLYDAAANLL----------TRPH-------TEAP 1208

Qy  1176  HTHPDMPDNIRLRKLKYDKAGRVINITDN------------HGNTEN---------FT 1209
          |                                   :   :::
Db  1209  HNQVQGSDKFDYRYDGFGRMVSRYEKGSSSGQRYHYDSDHRIIAVDIDQGPLGYQRAEYR 1268

Qy  1210  YDTIGRL------------QNGQGSVYGYDPLN-RLVSQKTDTLDCELYYRETM 1250
          |                  ||                         :
Db  1269  YDIILGRRIEKRLWKASAIANTVTYHQHEPDEVVTFGWVGMRLVSEHSSAAPHTTVYH--- 1325

Qy  1251  LVNEVRNGEMIRLLRTGETIIAQQRASKVLLTGSTDSQQSVILTSDKQNLSQRAYSAYG-- 1308
          |                  |              ||:         ::
Db  1326  AYNDQSYTPLARIECT-DNPLNPQRAIYTHSSLSGLPEALTNSEGEIVWQGQYSAWGHL 1384

Qy  1309  ----KHKSTANDASILGYNGERADPVSGVTHLGNGYRSYDPTLMRFHTPDSLSPFG-AGG 1363
          ||                                                     :|
Db  1385  QRQTRPTSTFNREONLRFQGQYFDKETGLHY--NTFRYXAPDLGRFTCQD---PIGLAGG 1439

Qy  1364  INPYSYCLGDPINRSDPSG 1382
          ||||: |  |       ||
Db  1440  INLYAYA-PNPLTWVDPWG 1457
```

RESULT 7
US-09-543-681A-7560
; Sequence 7560, Application US/09543681A
; Patent No. 6605709

```
Qy    768 TATKQLQSNSNNVQTGKEVTTYTPSQQPIQTLFDEAGHLQSCHTLTRDGWDRVRKETDA    827
              |||||
Db    860 LIBEQLPDGS----------------VRRYRYDTLGRQITRQDESGAI              891

Qy    828 IGQCTIYQYDNYNRVIQITLPDGTIVNRKYAPFSTDTLITDIRVNGISLGQQTFDGLSRL   887
              ||||||                              |||
Db    892 ------TRFQWDAAGRLSQITLPGG--ASRMY-------RYN--AYGKVT-           924

Qy    888 TQSQDDGGRVWAYTYSAGNDQCPSTVITPDGQFIHYQVQPELDDAVLQVASNEITQQFSYN   947
              :::::::::|:|:::::::|:||::
Db    925 SECDEQGRMTRYEY-LDDLHLVSRRINPDGSQLRYRY--ENARLLSEIENERGERYRLD    981

Qy    948 PVTGALLKAVAEGQSLTPIYPSGRLKMENINDMKKMSYLWTLRG--LE------NGYTDL   1000
                                      ::
Db    982 ------YHGNGLLISQETGFDGRRTAYRYDLKGQLLEKTEYGDDGSEL             1022

Qy   1001 TGTIQKISR---------------------DTHGRVTQIKDSSIKTTLNYDDLNRHIGSQVTD   1042
              |||||
Db   1023 RTTYQRDSTGRLLAKTLPDGNRVDYRYDTLGRLVAVDDGTWPLAYEYDLRDR--        1074

Qy   1043 LATGHMLTTTVB--FDGLNREIGRKLCD----------SSGHTL--DIQQSWLKTQQLANR   1089
              :::
Db   1075 LVREHQGWATLHYAYDALGQLIHCRLPDGNRVDYRYQTGGTLSAIDLNGQSLTRHQFGSG   1134

Qy   1090 IVKLNGVLQRTEQYSYDSRNRLNQYKCDGAECPTDKYGHSIVTQNFTYDIYGNITACHTT   1149
              ||||||
Db   1135 RERQRQGELLSQYHYDEQGRLLAHQVSQRQ---------RHLYQRQRYDASGNLAA-      1183

Qy   1150 FADGTEDHATFKFANPTDPCQLTEVHHTHPDMPDNIRLKYDKAGRVIN-----         1197
                                                            :
Db   1184 IEDSRKGIRSFHY-DPLD--RLLGVRGETPE-------SFVHDPAGNLLAQGGQPDARQMEV 1235

Qy   1198 ------IT----DNHGN---------------TENFTYDTLGR---LQNGQGSVVGYD    1227
              ||
Db   1236 RGNRLLTQGDRHPDYDAHGNLVRERRGTGQKLVTEYSYDCQHRLIGVSLPDGRQVVVRYD   1295

Qy   1228 PLNRLVSQKTDTLDCELYYRETMLVNEVRNGEMIRLLRTGETIIAQ--QRASKVLLTGTD   1285
              |||||
Db   1296 AFGRRIAKQVDG----------------RNTEFLWL---GERLLAESGDRHYRTYLYEPD   1336

Qy   1286 SQQSVLTS----------------DKQNLSQE-----------AYSAYGK--HKSTANDA   1317
              |||||
Db   1337 SFRPLALLDGBGPEQVEPCYQLDHLGTPQELTRADGRLCWSARYRAYGNVLKLDIAEVD   1396

Qy   1318 SILGYNGERADPVSGVTHLGNGYRSYDPTLMRFHTPDSLSPPGAGGINPYSYCLGDPINR  1377
              ||||||
Db   1397 NPLRFQGQYFDAETGLHY---NRHRYYNPSTGRYLTPDPIRL--AGGLNSYRY-VPNPTGW  1451

Qy   1378 SDPSG                                                        1382
              ||||
Db   1452 VDPLG                                                        1456
```

RESULT 9
```
; US-09-328-352-5542
; Sequence 5542, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
;  TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC99-03PA
;  CURRENT APPLICATION NUMBER: US/09/328,352
;  CURRENT FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5542
;  LENGTH: 1596
;  TYPE: PRT
;  ORGANISM: Acinetobacter baumannii
; US-09-328-352-5542

  Query Match             3.2%;  Score 285.5;  DB 4;  Length 1596;
```

```
Db    970 ----------RTQGELTQYRQYDKLGRTISTFSSSRDKQHPLNGITLWRKWFYDPQ--   1021
              ||:|:|                           |
Qy    970 DLHREIS------                                                 827

Db   1088 NRIVKLNGVLQRTEQYSYDSRNRLN----------------QYKCDGAECPT          1123
              ||||
Qy   1088 ...

Db   1022 GNLGAMEDTYRGWVEYLYDSEQRLKKVASSENLDAMLFYDRADNLLERPQSEMDAEHSPT  1081
              ||||
Qy   1022 ...

Db   1124 DKYGHSIVTQNFYDIYGNITACHTTFADGTEDHATFKFANPTDPCQLTEVHH          1176
              ||||
Qy   1124 ...

Db   1082 LELSPQGDKL-RQFGQWHYQYDAYGNVIA------RRYRN------QSSQTY-        1120
              ::|
Qy   1082 ...

Db   1177 THPDMPDNIRLKYDKAGRVINITDHGNTENFTYDTLGR                        1215
              ||||
Qy   1177 ...

Db   1121 ------------AYDGDNRLV-IAHNQGIKAQYHYDALGRRIHKTVENRESGQVKRQETHF 1168
              :|:
Qy   1121 ...

Db   1216 ----------LQNGQGSVY--------GYDPLNRLVSQKTDTLDCELYYRETMLVN----  1253
              :::
Qy   1216 ...

Db   1169 IWQGLRLLQEQDINTGKHQTYCYEHGSYTPLAVIVKQSSGF----HYWWHCDINSAPL   1224
              |||
Qy   1169 ...

Db   1254 EVRNGEMIRLLRTGETIIAQQ----------RASKVLLTGTDSQQSVILTSDKQONLSQEAYSA  1306
              |||||
Qy   1254 ...

Db   1225 EVTNAQ--------GNTIWSGKYERFGFVRSSPLSFSYSSPER----VMASFEQNLR----   1268
              :||
Qy   1225 ...

Db   1307 YGKHKSTANDASILGYNGERADPVSGVTHLGNGYRSYDPTLMRFHTPDSLSPFG-AGGIN   1365
              ||||||
Qy   1307 ...

Db   1269 ----------YAGQYFDNETGL-HF-NTFRFYDPQIGRFIMPD---PIGLLGGIN       1308
              ||||
Qy   1269 ...

Db   1366 PYSYCLGDPINRSDPSG-----------HLSWQAWTGI                        1392
              :::
Qy   1366 ...

Db   1309 LYQYA-PNPLGWIDPWGLSSQEMVRVRHHTSVEGLEGI                        1345
```

RESULT 8
```
; US-09-252-991A-23805
; Sequence 23805, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;  APPLICANT: Marc J. Rubenfield et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;  TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 107196.136
;  CURRENT APPLICATION NUMBER: US/09/252,991A
;  CURRENT FILING DATE: 1999-02-18
;  PRIOR APPLICATION NUMBER: US 60/074,788
;  PRIOR FILING DATE: 1998-02-18
;  PRIOR APPLICATION NUMBER: US 60/094,190
;  PRIOR FILING DATE: 1998-07-27
;  NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23805
;  LENGTH: 1626
;  TYPE: PRT
;  ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-23805

  Query Match             3.3%;  Score 292;  DB 4;  Length 1626;
  Best Local Similarity   21.1%;
  Matches  204;  Conservative  125;  Mismatches  316;  Indels 320;  Gaps  46;
```

```
Qy    597 HRSQVRSRY----TGRLFSDTDTKDIVTQMSYDKLGRLLTRTLNSGTPY----         641
              ::
Db    633 HTTQVLARYRCNERQLIEAGNALYEAERYRYDARNVILRRQLAGGASFWEWEGBGKQA    692

Qy    642 ----ANTLTYD-YELNN-----------LQDDNRPPFVITTTDVNGNQLRNE         677
              :|:
Db    693 RAVHHWASFPQMDSRYVWNEDGSVTAINADGSEEVYVHDDNAR--LVRQVPDPDGGETLRH   750

Qy    678 FDGAGRHVSQCLKDSDGDGKFYTIHTQQYDEQGR---------HHTSTYS-          718
              :|||
Db    751 YDEKGQLVAB----RDPLG-----TITEYRYDQAGRLEALLPAEGPTCYSYFDGFVRSVRR   803

Qy    719 -----DYLTNGR----QQTQPD--KVHLSMSKSYDNWGQIANTHWSYGVSEKITVDPITL   767
              :|:
Db    804 GEAQWKYERNAQGDITRQTDPEGNVTHY----AYDHRGCLVEITHADGSLHQLTWNPLGQ   859
```

  Query Match   3.2%; Score 285.5; DB 4; Length 1596;

Db 1194 GYDVFGRRIJYKKTAKELTLFGWDGDLMIWESFKSAQTNYTKHYIYEPDSFVPLLQAGYKD 1253

Qy 1343 YDPTLMRFHTPDSLSPFGAGGINPYSYCLGGDPI-NRS 1378
              : || :   :: | || || | ||: ||:
Db 1254 F---IQLIETPD---YQEYQTKPYS-IYKDPVNNRN 1282

RESULT 10
US-09-252-991A-23640
; Sequence 23640, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;   APPLICANT: Marc J. Rubenfield et al.
;   TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;   TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;   FILE REFERENCE: 107196.136
;   CURRENT APPLICATION NUMBER: US/09/252,991A
;   CURRENT FILING DATE: 1999-02-18
;   PRIOR APPLICATION NUMBER: US 60/074,788
;   PRIOR FILING DATE: 1998-02-18
;   PRIOR APPLICATION NUMBER: US 60/094,190
;   PRIOR FILING DATE: 1998-07-27
;   NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23640
;   LENGTH: 974
;   TYPE: PRT
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23640

Query Match   3.1%; Score 277.5; DB 4; Length 974;
Best Local Similarity 20.4%; Pred. No. 3.5e-12;
Matches 218; Conservative 141; Mismatches 377; Indels 335; Gaps 48;

Qy   504 EKFIQYRYSLIGSQSHVTLKIEERHYSATQLLNSTLFQYNTDKSELGRLLKQTECTKGEN 563
             || ||:   : || |:::||||: || |: : ||
Db   112 ERAWSYSYNALG---------LIKRADGPRTDVQDVTLYAYD------------SRGNL 149

Qy   564 GKTYSVVHKFTYTKQDDTLQQSHSITTHDNFTIHRSQVRSRYTG--RLFSDTDTKDIVTQ 621
              |:||| ||: | :: : | | : ::|:|| | :: | |
Db   150 TQVTNALGQVTRLGDYDERGKPGSITDANGVTSSLA-----YTGVDGWLASVSTAGSTTR 204

Qy   622 MSYDKLGRLLTRTLNSGTPYANTLJTYDYELNNLQDDNRPPFVITTTDVNGNQLRNEFDGA 681
              |: | : : |:   : | :   ::::|| | | ||
Db   205 FDYDAVGQITRVTRGDG----SWLSYEY-----DDARR--LVAIGNNLGERLEYDVDTK 252

Qy   682 GRHVSQCLKDSDGDGKFYTIHTQQ--YDEQGR------------HHTSTYS-DYLTNGRQQTDP 730
             ||:|| |||:|| |: | :: |: :|| : :|| | |
Db   253 GNRTAQRIKDASGS----LVRQQOWAYDELGRLLRAVGAGGQTRSFAYDLNDNPVGETNP 308

Qy   731 DKVHLSMSKSYDNWGQIANTHWSYGVSEKITVDPITLTATKQLQSNSNNVQTGKEVTTYT 790
              ::| || | | | | || | || | | | | ::   : | |: |::|
Db   309 RQ--FAHSQAFDALDRLVGQSDPLGGKTRLAYD------AQDNLTEVKDPRGVTTRY 357

Qy   791 PSQQPIQITLFDEAGHL----------QSCHTLTRDGWDRVRKETDAIGQCTTIYQYDNYNRVI 843
               : || :| | || | | | | |: | | || | :|
Db   358 E--------YDGLGNLTRLVSPDSGTTFEHDAAGNVIRRTDARGAVTEYRYDALNRLV 408

Qy   844 QITLPDGTIVNRKYAPFSTDTLITDIRVNGISLGQQTFDGLSRLTQSQDGGRVWAYTYSA 903
              | | || | : :|| ||: || |:: |: | |:| || :|
Db   409 ERRSPSDPSLDVQYR-------YDLTADGNK---------GIGRL-GAIDGAR------ 444

Qy   904 GNDQCPSTVITPDGQFIHYQYQPELDDAVLQVASNEITQQFSYNPVTGALLKAVABGQSL 963
             || | |:| || :||: : :| :: : | :: || |: ||| |
Db   445 --------DSLVYRYD-ERGNLVEQVRSIRLDQQTLLDRVTYRYDAA-----NQL 485

Qy   964 TPIYPSGRLKMENINDMKKMSYLMTLRGLENGYTDLTGTQIKSRDTHGRVTQIKDSSI 1023
             | |:|| ||| ||| | :||:| | ||||||| : :| || || | ||:|
Db   486 ----------------------GLAIGY-----------------PRNAGGQVASV 509

Qy 1024 KTTLNYDDLNRHIGSQVTDLATGHMLTTTVEFDGLNREIGRKLCDSSGHTLDIQQSWLKT 1083
              : |:| : | | || :|| || || | |||| ||| || | ||| |
Db   510 -TLAVGDKAPSTLVGQIAYLPFGPLLRLT---------------WGNGITLSREYDQD 551

— Left column —

Best Local Similarity 18.1%; Pred. No. 2e-12;
Matches 202; Conservative 179; Mismatches 387; Indels 349; Gaps 49;

Qy   440 ESGNPLTKVIKDKKTQKIISPSTHWEYYPPAGEVDNCPEPYGFTRFVKKIIQTPYDSEF 499
             :|:| |: :|| | | ||: | | | | :   : |: :
Db   337 QSGKSISYSIGABRVQ--------HADFVLP----KIGFS-FIRQ----YNSQM 373

Qy   500 KDDPEKFIQYRYSL------IGSQSHVTLKIEERHYSATQLLNSTLFQYNTDKSELGRL 552
             |:| || |||| |||:| |:|||||||: | |||||||| ||||||||
Db   374 DEFDQSMVGARWMPFSNMIQQNAQGYLFIDSKGKRKH--QLPVSIIFETYEVPYE-GWI 429

Qy   553 LK------------------------QTECTKGENGKTYSVVHKFTYTKQDDTLQQSHSITTH- 591
             |:   :: || || |:|| ||:||| | : : |
Db   430 IKPLKNGELILDFGGEWRSHFQSFDGGKNYLVKKMNETSQEILLE-YLLLDHIAYLKV 488

Qy   592 DNFTIHRSQVRSRYTGRLLFSDTDTKDIVTQMSYDKLGRLLTRTLNSGTPYANTLTYDYEL 651
             :| |: | :|| : | : |: ||:
Db   489 INFKLKQAEYELKP--------AFNEQVKIIAVPL----------------- 515

Qy   652 NNLQDDNRPPFVITTTDVNGNQLRNEFDGAGRHVSQCLKDSDGDGKFYTIHTQQYDEQGR 711
              |:|:||| : | ||: |: || | | |::|
Db   489 LK-----QTECTKGENGKTYSVVHKFTYTKQDDTLQQSHSITTH----- 591

Qy   712 HHTSTYSDYLTNGRQQTDPDKVHLSMSKSYDNWGQIANTHWSYGVSEKITVDPITLTATK 771
             | | : |||| |::   : | | | || | | : ||:|| |||
Db   516 ---RYEYDTQG-----NLIKAIDQNG-----HTRTYEYNQF 551

Qy   772 QLQSNSNNVQTGKEVTTYTTPSQQPIQITLFDEAGHLQSCHT----------LTRDGWDRVRKE 824
             |||:|||| |||:|| |:| :|||||||||| | |:| | | ||| ||
Db   552 HQLTRYTD--RTGRGQNIRYESTEAKAKAIEEWADDGSFH------TKLKWHP---RLR 599

Qy   600 QVA-----VYDAYDVPTV----------YYFDLDGFTYRTRLADGRESWYSRDGKKKRITRQ 645
              |: | |:| | | ::| | |:   :: |
Db   825 TDAIGQQCTIYQYDNYNRVIQITLPDGTIVNRKYAPFSTDTLITDIRVNGISLGQQTFDGL 884

Qy   646 IDFDGRETQQEYNDQDQLVKIVQPNGGIIRFAY 681
             : || :| |::||:   :: || ||| |
Db   885 SRLTQSQD-GGRWAYTYSAGNDQCPSTVITPDGQFIHYQYQ-----PELDDA----VLQV 935

Qy   682 GNLVEIKDPEGSIWKREYDENRN--VSKEINPLGHITQYKYNNDNQLVEVIDAKGGVKKI 739
             || || | | : || :| | | | || |: |::|:||: | | | |
Db   936 QYNELGQMISYTD--------CSGKSSTWEYDEDGALTAEQTANNKVVQYFYSTKGRDK 995

Qy   740 ASNEITQQPSYNPVTGALLKAVABGQSLTPIYYPSGRLKMENINDMKKMSYLWTLRGLEN 799
             ||:| || :|||| || ||:|:| ||: | | |::|:| |:|
Db   996 YT----------DLTGTIQKISRDTHGRVTQIKDSSIKTTLNYDDLNRH-------- 1035

Qy   800 GYT------------------------DLTGTIQKISRDTHGRVTQIKDSSIKTTLNYDDLNRH 849
             |:   :|| |::| | |::|| | |: | |:|| | |
Db 1036 ---------IGSQVTDLATGH----MLTTTVEFDGLNREIGRKLCDSSGHTLDIQQSWL 1081

Qy   850 WDKQGRIQKLINQNQAEYLFGYNPYGVLFREQAFDGEEKHYS--YNENGRLFQIRRPNI 906
             | ||| || ||:|:|| | |:||:| |:: :|| |||: | |||| |:| |:
Db 1082 KTQ-----QLANR-IVKLNGVLQRTEQYSYDSNRNLNQYKCDGAECPTDKYGHSI--- 1130

Qy   907 LTQFDYYADGQIASKSFTHLHTGQKQTEQFDYNLNSQLSRASNEVSQ--IDLYRNALGQL 964
             ||| |:|| :|:|||| | |||:: |:| | || ||:|| |:| | | | |:|
Db 1131 ------------VTQNFTYDIGNI-----TACHTTFADG-----TEDH 1157

Qy   965 VREHQHYKIPELKPLTAVLHYEYDELGNLIKTIRPDGHTLNHLVYGSGHIYAIGLNNQEV 1024
             | |||| || |||:|| | |:|:|:||:|| : ||| || ||| | ||:|||||:|
Db 1158 ATF---------KFANPTDPCQLTEVHHTHPDMPDNIR-----K 1186

Qy 1025 VSFQRDDLHRETRLLANGLMQTKQYNDVGLLSSQFNQPEQETQDYLQYQAHR---------K 1078
             |||:|| | | || | | |:: || :::|:| || |: | :|: |
Db 1187 LKYDKAGRVINITDNHGNTENFTYDTLGRLNYQDPIGRLIAAQSLHKTESFNFDPAGNLIDSBSSVLSPA 1242

Qy 1079 YHYDKNYLLSQVEDSRLGKLNYQDPIGRLIAAQSLHKTESFNFDPAGNLIDSBSSVLSPA 1138
             | ||| | ||| | ||:||| |: | | :|:|:| | |
Db 1243 ELYRETMLVNEVRNGEMIRLLRTGETIIAQQRASKVLLTGTDSQQSVILTSDKQNLSQB 1302

Qy 1139 QI---KNNLIKSYKGKHYQYDVQGNVTEIIQ--AGKNLKLTWDNQNRLIRSDNNGLVTEY 1193
             |: || ||: ||: | ||: |:|: ||:| | | |:| |:|:|| | |
Db 1303 AYSAYGK--HKSTANDASILGYNGERA------------DPVSGVTHLGNGYRS 1342

```
Qy  700 TIHTQQYDEQGR--HHTSTYSDYLTNGRQQTDPDKVHLSMSKSYDNWGQIANTHWSGYVS 757
                  :  :::  ::::: ::
Db  807 -----DYNDYGQLTHFTSAFGD------------EWQLSYDENGNLT- 836

Qy  758 EKITVDPITLTATKQLQSNSNNVQTGKEVTT------YTPSQQPIQITLFDEAGHLQS 809
                                                    :::
Db  837 --IVTDP---QGRQQVYEYSQHGELLKAITPNGAQWQYHYNPAHQLIKTT--NPYQHSTE 889

Qy  810 CHTLTRDGWDRVRKETDAIGQCTIYQ------------YDNYNRVI 843
            ::  :    ::                          ::::::
Db  890 YHS---DELGRLLHYTDALNHTTRYQYSTEHASTNGSLSKILLPDGVEQQIDYDSERRVI 946

Qy  844 QITLPDGTIVNRKYAPFSTDTLITDIRVNGISLGQQTFDGLSRLTQSQD-GGRVWAYTV- 901
            ::::  ::          ::   :::::::
Db  947 AVTDGEGKTTRYRYGPF--DVLLAMIRPDGSEI-RFEYDSLTRLKKVVNANGEVYLYERD 1003

Qy  902 SAGN----------DQCPSTVIT--PDGQFIHYQYQPELDDAVLQVASNE----- 939
          :    :::
Db 1004 KAGQIIREVDFTGREICYRYDRLGRRIATRYPDNHELRWRY---NESGLVVEQSEWFED 1059

Qy  940 ------ITQQFSYNPVTGALLKAVAEGQSLTPIYYPSGRLKMENINDMKKMSYLWTLRG 992
               :  :::  ::                 :  ::
Db 1060 EQESRCLSTAQYSYN-ARQQLIKATNPDSVVEFEYDDQGRLCSERINE-QEIVHQM--- 1113

Qy  993 LENGYTDLTGTIQKI-SRDTH------GRVTQIKDSSIKTTLNYDDLNRHIGSQVTDLAT 1045
           ::                          :
Db 1114 --NEADNTLALTRFGERELHYAFGALGELTSLQ---------VNQHAPLQFS---- 1154

Qy 1046 GHMLTTTVEPFDGLNREIGRK---LCDSSGHT-----LDIQQSWLKTQQLANRIVKLNGVL 1097
                       :::                         ::
Db 1155 YNAVGQEYLRRSRAGFVNSSHYTATGLLAHQRAGRGTEQF------LQSIQ 1199

Qy 1098 QRTEQ--------YSYDSNRLNQYKCDGAECPTDKYGHSIVTQNFTYDIYGNITACH 1147
                         :::                  :::    ::
Db 1200 AHPQQPPFCTDVHRSYQYD----RAYNVVGIE--DDRWRQT--RYHYNANDQITETQ 1248

Qy 1148 TTFADGTEDHATFKFANPTDPCQLTEVHHTHPD--MPDNIR------LKYDKAGRVINI 1198
                                                            ::: ::
Db 1249 YSPQWGNQDE-KFQYDNNLN---ITE-HLTTPSSSNVPSDAQGAMLQLFQQQQAGRVTRR 1303

Qy 1199 TDNHGNTENFTYDTLGRLQ---------NGQGSV------YGYDP 1228
              :      ::
Db 1304 YTAKG-YQDYHYDVNGRLAKKIVHTRGFRPREWRYLMNTQNQLTACFTPKGDCWHYTYDA 1362

Qy 1229 LNRLVSQKTDTLDCELYYRETMLVNEVRNGEMIRLLRTGETII---------AQQ- 1274
                 :  ::                  :::
Db 1363 FGRRLS-KTKTVDSDLAHIDPLFPQIKPKITTWRYLMSGDQLIEETPIYADGTLANAQQV 1421

Qy 1275 ---RASKVLLTGTDSQQSV-ILTSDKQNLSQEAYSAYG-------------KH 1310
              ::
Db 1422 QWLYQPGEITPFTARYQQGKLHYYVVTDHQGTPREIFSFSEGGQASWAGRLNTWGQMQPFWRYRD 1481

Qy 1311 KSTANDASI----LGYNGERADPVSGVTHLGNGYRSYDPTLMRFHTPDSLSPFG-AGGIN 1365
                          ::                   :::
Db 1482 GKAENDPNYTECPFRFAGQYEBDEESGLYY--NRFRYYDRETGQYLSPD---PIGLLGGLN 1536

Qy 1366 PYSY--CLGDPINRSDPSGHLSWQAWTGIGMGIAG----LLLTIAT 1405
          :::                                    ::::::
Db 1537 PYGYVHC---PTGWVDP--------FGLAGDCCDELLVLAT 1566
```

RESULT 12
US-08-851-567B-61
; Sequence 61, Application US/08851567B
; Patent No. 6528484
; GENERAL INFORMATION:
; APPLICANT: Ensign, Jerald C
; APPLICANT: Bowen, David J
; APPLICANT: Petell, James
; APPLICANT: Fatig, Raymond
; APPLICANT: Schoonover, Sue
; APPLICANT: ffrench-Constant, Richard
; APPLICANT: Rocheleau, Thomas A.

```
Qy 1084 QQLANRIV----------KLNGVLQR--------TEQYSYDSRNRLNQYK-CDGAECPTDKY 1126
           :                                  :::
Db  552 YQLLRQKVGPWQSDYQHDANCNIQQHRHSLWGTLDYQYDPLDRLTEERGVQGG------ 604

Qy 1127 GHSIVTQNFTYDIYGNIT-ACHTTFADGTEDHATFKFANPTDPCQLTEVHHTHPDMPDNI 1185
                          :::  ::  ::::
Db  605 ------RSYAYDAVGNRTQRSDNPASGGTASSQDYQYA------------PDSN 640

Qy 1186 RL----------KYDKAGRVINITDNHGNTENFTYDTLGRLQN-----GQGSVYGYDPL-NR 1231
          ::          ::: ::
Db  641 RLTAIGAQAVTSDAAG---NLTQDRA-ARKLAYDAQGRLQSVSLDGQQVAEYRYNALGQR 696

Qy 1232 LVSQKTDTLDCELYYRETMLVNEVRNGEMIRLLRTG------ETIIAQQRAS---- 1277
              :  ::                  :::
Db  697 IVKLTPESVTTYLYGPDGQLLGEAEHDGSGRKLRAQYYLWLDSLPLATIDADYDAQGKVG 756

Qy 1278 --KVLLTGTDSQQSVILTSDK--QNLSQEAYSAYGKHKSTANDAS--ILGYNGERADPVS 1331
              ::                 :::
Db  757 NPTLLYLHGDHLDTPRLATDASGQIAWQWQSDAFGRGEALSQGSTQVNLRFPGGYYDAES 816

Qy 1332 GVTHLGNGYRSYDP-----------TLMRFHTPD-----------SLSPFGAGGINP-- 1366
          :::                      ::::
Db  817 GLRY--NYPRDYDPETGRYVESDPVETLRKLNNPEMTFLNSGESMLQATPYWEHGFTPNH 874

Qy 1367 -YSYCLGDPINRSD-----------PSGHLSW---------QAWTGIGMGIAGLLLTIAT 1405
           :::                     ::::
Db  875 NYTYSDNNPTAKSDKHGLSPNPTDNLIYTPDTNCTCTLECKKKFTGNGKSF------ 925

Qy 1406 GGMAIAAAGGIAAAIASTSTTALAFGALSVTSDITSIVSGALEDASPKASS 1456
          :::
Db  926 ------LVGALCSKATTPF-FGGVVCNSTIVMICGASCSQECNRAPS 965
```

RESULT 11
US-09-543-681A-4476
; Sequence 4476, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABIL
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 4476
; LENGTH: 1665
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-4476

Query Match   3.1%;   Score 271.5;   DB 4;   Length 1665;
Best Local Similarity 20.7%;   Pred. No. 2.6e-11;
Matches 233; Conservative 148; Mismatches 350; Indels 395; Gaps 66;

```
Qy  508 QYRYSLIGSQSHVTLKIEERHYSATQLLNSTLFQYNTDKSBELGRLLKQT-----EC 558
                                               ::  ::
Db  608 QHRIDFIYNEQSQLVSV-KRHDGIALIL------BYQQDK--LVRICSESTTSSPRDFVRC 659

Qy  559 TKGENGKTYSVVHKFTYTKQDDTLQQSHSITTHDNFTIHRSQVRSRYTGR-----LFSDTD 614
            ::  ::
Db  660 EYDTHG--------YLSQCHAYQQNH-------LWHRYSPEGLMVAWGDTD 695

Qy  615 TKDIVTQMSYDKLGRLLTRTLNSG-------TPYANTLTY-------DYELNNLQDDNRPP 661
              ::  ::
Db  696 STELT--IDYDEQGRVVATHSPSGFWNDRFIYDDYQRMTYIDABGGFSRYYYNDDN--- 750

Qy  662 FVITTT-------------DVNGNQLRNEFDGAGRHVSQCLKD--SDGDGKFY 699
                                    ::  ::
Db  751 -LVTRTIDPLWRETYTEWEQRKKIAEINEIGERTEY---GYHVNGLLAYILYPDGKAIYY 806
```

```
;   APPLICANT:    Blackburn, Michael B.
;   APPLICANT:    Hey, Timothy D.
;   APPLICANT:    Merlo, Donald J.
;   APPLICANT:    Orr, Gregory L.
;   APPLICANT:    Roberts, Jean L.
;   APPLICANT:    Strickland, James A.
;   APPLICANT:    Guo, Lining
;   APPLICANT:    Sukhapinda, Kitisri
;   APPLICANT:    Ciche, Todd A.
;   TITLE OF INVENTION:  Insecticidal Protein Toxins From Photorhabdus
;   NUMBER OF SEQUENCES:  88
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Dow AgroSciences Patent Department
;     STREET:     9330 Zionsville Road
;     CITY:       Indianapolis
;     STATE:      IN
;     COUNTRY:    US
;     ZIP:        46268
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/851,567B
;     FILING DATE:  05-MAY-1997
;     CLASSIFICATION:  514
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 08/063,615
;     FILING DATE:  18-MAY-1993
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 08/395,497
;     FILING DATE:  28-FEB-1995
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 60/007,255
;     FILING DATE:  06-NOV-1995
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 08/608,423
;     FILING DATE:  28-FEB-1996
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US 08/705,484
;     FILING DATE:  28-AUG-1996
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Seay, Nicholas J
;     REGISTRATION NUMBER:  27386
;     REFERENCE/DOCKET NUMBER:  960296.93804
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  608-251-5000
;     TELEFAX:  608-251-9166
;     TELEPHONE:  608-251-9166
;   INFORMATION FOR SEQ ID NO:  61:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1043 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-851-567B-61
```

```
Query Match     2.7%;   Score 238.5;  DB 4;  Length 1043;
Best Local Similarity  20.4%;  Pred. No. 4.2e-09;
Matches  221;  Conservative 130;  Mismatches 378;  Indels 353;  Gaps  51;

Qy   612  DTDTKDIVTQMSYDKLGRILTRTLNSGTPYANTLTYD---YELNNLQDDNRPPFVITTTD  668
          :  |::|  :|:  :|:|
Db    38  DTDTR--VTRHQYDARGHL--------------------NYSIDPRLYDAKQADNSVKPNFV-WQHD   81

Qy   669  VNGNQLRNEFDGAGRHVSQCLKDSDG------------------------DGKFYTIHTQQY  706
          :  |:|:  |::|  :|:|  :|  |  |  :|  :|
Db    82  LAGHALRTESVDAGRTVA--LNDIEGRSVMTMNATGVRQTRRYEGNTLPGRLLSVSEQVF  139

Qy   707  DEQGRH--------HTSTYSDYLTNGR--QQTDPDKVHLSMSKSY----------------  741
          :::           :|:|:  :|:|::|
Db   140  NQESAKVTERFIWAGNTISEKEYNLSGLCIRHYDTAGVTRLMSQSLAGAMLSQSHQLLAE  199
```

```
Qy   742  ---DNWGQIANTHWSYGVSEKITVDPITLTATKQLQSNSNNVQTGKEVTYTPSQQPIQI  798
            ||                     ||             :::
Db   200  GQEANWSGDDETVWQGMLASEVYT----------------TQSTTNAIGA---LLTQTDAKGNIQR  246

Qy   799  TLFDEAGHLQSCHTLTRDGWDRVRKETD-----------AIGQCTIYQYDNYNRVIQITLP  848
          ||        ::  ::
Db   247  LAYDIAGQL------KGSWITVKQSEQVIVKSLSWSAAGHKLREEHGN------------  289

Qy   849  DGTIVNRKYAPFSTDTLITDIRVNGISLGQQTFDGLSRLTQSQDGGRVWAYTYSAGNDQC  908
          |::|                  ||
Db   290  -GVVTBYSYEP-ETQRLI-------GIT------TRRAEGSQSGARVL-------------  322

Qy   909  PSTVITPDGQFIHYQYQPELDDAVLQVASNEITQQFSYNPVTGALLKAVAEGQ----SLT  964
                                   |: |:|
Db   323  --------QDLRYKYDPV--GNVISIHNDAEAATRFWRNQ---KVEPENRYYVDSLY  365

Qy   965  PIYYPSGRLKMENINDMKKM--SYLWTLRGLENGYTDLTGTIQKISRDTHGRVTQIKDSS  1022
          :                     |:||:
Db   366  QLMSATGR-EMANIGQQSNQLPSPVIPVPTDDSTYTNY--LRTYTYDRGGNLVQIRHSS  421

Qy   1023 IKTTLNY-DDLNRHIGSQVTDLATGHMLTTTVEFPDGLNRREIGRKLCDSSGHTLDIQQSWL  1081
          ||:                  |::            |::
Db   422  PATQNSYTTDIT--VSSRNRAVLSTLTTDPTRVDA--------LFDSGGH----QKMLI  467

Qy   1082 KTQQLANRIVKLNGVLQRTEQYSYD--SRNRLNQYKCDG--AECPTDKYGHSIVTQNFT  1136
          |:  |                      :| :|    |::
Db   468  PGQNLDWNI---RGELQRVTPVSRENSSDSEWYRYSSDGMRLLKVSEQQTGNSTQVQRVT  524

Qy   1137 YDIYGNITACHTTFADGTEDHATFKPANPTDPCQLTEVHHTHPDMPDNIRLKYDKAGRVI  1196
                                           |:|:
Db   525  Y------------LPGLELRTTGVADKTTEDLQVITVGEAG---RAQVRVLHWESGKPT  568

Qy   1197 NITDNHGNTENFTYDTLGRLQNGGQGSVYGYDPLNRLVSQKTDTLDCELYRETMLVNEVR  1256
          :|:
Db   569  DIDNNQVR---------YSYD--NLLGSSQLE-------------------------  589

Qy   1257 NGEMIRLLRTGETIIAQQRASKVLLTGTDSQQSVILTSDKQNLSQEAYSAYG------KH  1310
                                           |::
Db   590  ---------------LDSEGQILSQEEYYPYGGTAIWAAR  614

Qy   1311 KSTANDASILGYNGERADPVSGVTHLGNGYRSYDPTLMRFHTPDSLSPFG-AGGINPYSY  1369
          |:               |:::
Db   615  NQTEASYKFIRYSGKERD-ATGLYYY--GYRYYQPWVGRWLSAD---PAGTVDGLNLYRM  668

Qy   1370 CLGDPINRSDPSG---------------HLSWQAWTGIGMGIAGLLLTIATG--GMAIA---AAG  1414
          |:                         |:|:
Db   669  VRNNPITLTDHDGLAPSPNRNRNTFWFASPFLFRKPDEGMSASMRRGQKIGRAIAGGIAIG  728

Qy   1415 GIAAAIASTSTTAL--AFGALSVTSDITSIVSGALEDASPKASSILGWVSMGM------  1465
          |::  |:  |:::
Db   729  GLAATIAATAGAAIPVILGVAAVGAGIGALMGYNVGSLLEKGGALLARLVQGKSTIVQSA  788

Qy   1466 --GAAGLAESAIKG----GTKLATHLGAFAEDGENALLKSTSESSRIKWGVTRSLDREIV  1519
            |:|:  |:  |:
Db   789  AGARAAGASSAAAYGARAQGVGVASAAGAVT-----------------GAVGSWINNAD  829

Qy   1520 RNEEGQVIKDHSRGYTDNFMGKGBQAILVH------------GDKDGFLYHTEGNK----  1568
          |:|:                                     |::|
Db   830  RGIGGAIAGSAVGTIDTMLGTA--STLTHEVGAAAGAAGGMITGTQGSTRAGIHAGIG  887

Qy   1569  PY 1570
          |
Db   888  TY 889
```

```
RESULT 13
US-09-328-352-5545
; Sequence 5545, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
```

```
;  FILE REFERENCE: GTC99-03PA
;  CURRENT APPLICATION NUMBER: US/09/328,352
;  CURRENT FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 8252
;  SEQ ID NO 5545
;  LENGTH: 804
;  TYPE: PRT
;  ORGANISM: Acinetobacter baumannii
US-09-328-352-5545

Query Match      2.7%;  Score 235.5; DB 4;  Length 804;
Best Local Similarity   20.8%;  Pred. No. 4.6e-09;
Matches 147; Conservative 92; Mismatches 253; Indels 215; Gaps 30;

Qy   699 YTIHTQQYDEQGRHHTSYSYLTNGRQQTDPDKVHLSMSKSYDNWGQIANTHWS------ 753
            :|:: |||:::: : : :|:  :|  ::  :::    ::|:  ::
Db    21 FTLQTTEYDAQGAIRSNTQAGRVTKYKYET-----VSFSTCL--GNLASFQWSKFVSC 71

Qy   754 ---YGVSEKITVDPITLTATKQLQSNSNNVQTGKEVTTYTPSQQ-PIQITLFDEAGHLQ 808
            ::  :   :| :|:    :::    ::||: |||  ::  :
Db    72 LSNYVLGDQHVGLPIEV----------DVGNGKEITQPAGYKRGIPTQVKLANGA--- 116

Qy   809 SCHTLTRDGWDRVRKETDAIGQCTIYQYDNYNRVIQITLPDGTIVNRKYAPFSTDTLITD 868
            : : |  : :  :  ::   :|:::|:|:  :|:   : |: :|
Db   117 -TETNIVDDFGNITQHTDADGVISRKQYDDAGRL----YIDTPIVGLNYSTFTYDGLTVS 171

Qy   869 IRVNG----------------------IS---LGQQTFD--------GLSRLTQS- 890
            | :|                       :    |:  ::         | |  :
Db   172 RVVTGGGQLSRIEKYNGDGLLISSBDKISNKSIINSNKYDAFGNLIFKSNPGFSAITSGT 231

Qy   891 --------------QDGGRVWAYTYSAGNDQCPSTVITPD----------GQFIH---- 921
                          ::|  : : |: :|:|  ::|  :           | ::
Db   232 TSSYDVFDRPITVNDNGSVVTYCYQSCGGKTGAIVQTTDSFGTTESNLLAAGDFSADLKT 291

Qy   922 ----------YQYQPELDDAVLQ--VASNEITQQFSY---------NPVTGALLKA 956
                      : :|  |  :|:   :|::  | |::          |||||||:|
Db   292 LVARKGTDGSVFQTTEFENALLKPKVAVSGSSTQSYTYNSNTTLATEKDNSISGQKTFK 351

Qy   957 VAEGQSLTPIYYPSGRLKMENINDMKKMSYLWTLRGLENGYT-DLTGTIQKISR---- 1009
            |: ::      :   ||  || :: :::: : ||    ::   |:||:|||:||
Db   352 YDDTGRITSITHPDSSVETIKYFQLKDLIASRTWREVETTYSYSLAGRLKTTTNANISEA 411

Qy  1010 ---DTHGRV-----------TQIKDSSIKTTLNYDDLNRHIGSQVTDL--------A 1044
               |:|||            |:||:: |:   :|   :  ::| |  :         |
Db   412 FDLDTYGRVLISHTQKINANDTNNSYVVRYGYNQLN------QVTSIQYPNGKSVNLSNQNA 466

Qy  1045 TGHMLTTTVEFPDGLNREIGRKLCDSSGHTLDIQQSWLKTQQLANRIVKLNGVLQR---TE 1101
            |||  :|| :  :|:: ::   :  ||    : ||  : | :|   || | ||:   :|
Db   467 LGEVTSIPNVIQSLNYNARQQLTTVQANTDTLWSYTYNDSGLLNNISATS--LEKCVLNV 524

Qy  1102 QYSYDSRNRLNQYKCDGAECPTDKYGHSIVTQNFFYDIYGNITACHTTFADGTE-DHATF 1160
            :| |:: ||:|: | |::   :|:::|  ::  : | :||  ::| ::  ||||   :|
Db   525 DYGYDKLNRVNKL------SDKKCGS---VYNAYIDRYG------TGLMSTVELDQARY 567

Qy  1161 KFANPTDPCQLTEVHHTHPDMPDNIRLK-----------YDKAGRVINITDNHGNTENFTY 1210
            ::|:  | |:||  :   ||  ::::||            ||:||::| :|:: |:||  ||
Db   568 QYSYNND--DITKV--------NITSKSSTVAPAIYTNYVNNTSRLASVSGSTYSFFTY 616

Qy  1211 DTLGRLQNGQGQSVYGYDPLNRLVSQKTDTLDCELYYRETMLVNEVRN 1257
            | :||: |  |:  | ||  :||:   ||:   : |||:|::|:  |
Db   617 DAMGNVTNDGVRTLTYDNYSRLNKNGNETY---LYNADGLRVRAVRD 660
```

```
;  CURRENT FILING DATE: 1999-02-18
;  PRIOR APPLICATION NUMBER: US 60/074,788
;  PRIOR FILING DATE: 1998-02-18
;  PRIOR APPLICATION NUMBER: US 60/094,190
;  PRIOR FILING DATE: 1998-07-27
;  NUMBER OF SEQ ID NOS: 33142
;  SEQ ID NO 31032
;  LENGTH: 1128
;  TYPE: PRT
;  ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31032

Query Match      2.5%;  Score 223;  DB 4;  Length 1128;
Best Local Similarity   19.6%;  Pred. No. 7.8e-08;
Matches 258; Conservative 151; Mismatches 467; Indels 442; Gaps 68;

Qy   235 NFSLGNENPLTWSFGYTP----IGKNGILGQWITSMTAPGGLKETVNYSNNNQGHH--FP 288
            :|  |:||  :|| |:|     ||:||:||:|:|:| :|||:||  |  |::| ||  ||
Db    63 DFALAAPDTFTFARGYLSSNPRIGRLG-RGWWLPGESMHLELSEDACVLVDAQGRRIGFP 121

Qy   289 QSANLPVLPYVTLMKQVPGAGQPAIQAEYSYTSHNYVGGGSNGIWNNKLDNLYGLMTEYN 348
                      :|  :  |||||     : :|::||:||||||:||||:|||:|||||||||
Db   122 ALA----------PGA----QHYS--------GSSELW 137

Qy   349 YGSTESRRYKDKEGHDQIVRIERTYNNYHLLTSECKQQNGYIQTTETAYYAIIGHNFDSQ 408
               :|||:  :  ::|  |:| |  |   || :|  :    :  |::||  ::|::
Db   138 -----LRRGGSSGGEAQAWR-------LQTQEGSVLVLSGHSY-- 177

Qy   409 PSQFQLPKTKTETWRSADNSYRSEIT-------ETTPDESGNPLTKVIKDKK----TQ 455
             |:| :|  : :: | |::|  ::       ::||| :|||||:||:||||    ||
Db   178 -LHFQ--------RCPDGIWRLQASFGRAGYRTEFRWSGRGLLTGVRDSAGRSYALVYQ 227

Qy   456 KIISPSTHWEYYPPAGSVDNCPPEPYGFTRFVKKIIQTPYDSEFKD--DPEK----FIQ 508
              :  || :|    :   | : |: :||::::  :::   |::  :|  ||:|    ||:
Db   228 QACEPS---------EGDD------GLRLF--GVILASHDGPPPDYIDPQSPGLDWLVR 269

Qy   509 YRYSLIGSQSHVTLKIEERHYSATQLL------NSTLFQYNTD-KSELGRLLK 554
            |::| ||| :|||:|:||||:|||||:      |||||||||| |||||||||
Db   270 YQFSDSGD----LIAVRDRLGQVVRVFAWREHMLVAHGEPGGLEVRYEWDVHAPHGRVVK 325

Qy   555 QTECTKGENGKTYSVVHKFTVTKQDDTLQQSHSITTHDNFTIHRSQVRSRYTGTRLFSDTD 614
            |:|  :|| ||: ||:  : : :::   : |||:| :::| |||||:|||||||||:| ||
Db   326 QIEAG--------GLTRTFRYLR--DATEVSDSLGRVERYEPAGEGGQRRWTALVRADGS 375

Qy   615 TKDIVTQMSYDKLGRLLTRTLNSGTPYANTLTYDYELNNLQDDNRPPFVITTDVNGNQL 674
            :||:||||||||||||||||||||||||||||||||||:|||||||||||| |:||| |
Db   376 R-----SEFDYDLFGRL--------VAMRDPLGRET 398

Qy   675 RNEFDGAGRHVSQCLKDSDGDGKFYTIHTQQYDEQGRHHTSYSDYLTNGRQQTDPDK-- 732
            |  |||||||   | | |:|:||: |:|||||||||||||||||||||||||||||||
Db   399 RRRRDGQGRMLEE--ESPGKARY 429

Qy   733 -VHLSMSKSYDNWGQIANTHWSYGVSEKITVDPITLTATKQLQSNSNNVQ-TGKEVTTY 789
             |:||||||||||||||||||||||||||||||||||||||||||||:: | ||||||
Db   430 LVELE--------DAMQRRWTFERDER 465

Qy   790 TPSQQPIQIT-LFDEAGHLQSCHTLTRDGWDR---VRKETDAIGQCTIYQYDYNRVIQI 845
            |||||||||: |||||||||||||||||||||   |||||||||||||||||||||||:
Db   466 EDPRLPDRPTRIVDPRGGER-------RLEWNRFGLLAALTDCSGQVWRYDYDNEGRLVAS 519

Qy   846 TLPDGTIVNRKYAPFSTDTLITDIRVNGISLGQQTFDGLSRLTQSQDG-------G 894
            |:|  |:  | | |:  |::: ::|::| ::  :  : : |||:  |         |
Db   520 SDPLGQLTRRRYDPLG--QLIGLELADGSALSYE-YDALGRQTRIADAEBGHATLFSWGHG 576

Qy   895 RVWAYTYSAG-----NDQCPSTVITPDGQFIHYQYQPE-LDDAVLQVASNEITQQFSY 946
            | :|||:|||     |||||||||||||||||||||:  |||||||||||||||||||
Db   577 DLLARVSDAGGGELSYIHDEAGRLVALTNENGVQAQFRYDLLDRLVEETGFDGRRQRYRY 636

Qy   947 NPVTGALLKAVABGQSLTPIYYPSGRLKMENINDMKKM------SYLWTLRG---- 992
            ||||||||||| |||||||||||||||||||||||||||      |||||||||
Db   637 NAADELIAREDADGRETTYAYDRDGRLLASIRVPATEHAPALVERYRWLADGRLASACAGAD 696

Qy   993 LENGYT-DLTGTIQKISRDTH---GRVTQIKDS---SIKTTLNYDDLNRHIGSQVTDL-- 1043
```

```
RESULT 14
US-09-252-991A-31032
; Sequence 31032, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT:  Marc J. Rubenfield et al.
; TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION:  AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE:  107196.136
; CURRENT APPLICATION NUMBER:  US/09/252,991A
```

```
Db   697 CEVRYTYDEVGNL-RLESQVHADGWVYSVEHSHDALGVGRVQTSRYGD-----APPVAWLTY 750
                                                                    SMTAPGGLKETVNYSNNNQGHHFPQSAN
Qy  1044 ATGHM--LITTVEF---DGLNREIGRKLCDSSGHTLDIQQSWLKTQQ------LAN 1088
                                                                    BWTEIPABFLWSVEPQTNISIDISLTDQAGNTAIITHTQN--YNVDPTPN
Db   751 GPGHLHGALVGAVELAFERDALHREVRRDA------RRDGQDDALFTQERQHAPLGRLQR 804
                                                                    LPVLPVTLMKQVPGAGQPAIQAEYSYTSHNYVGGGSNGIWNNKLDNLYGLMTEYNYGST
Qy  1089 RIVKLNGVLQRTEQYSYDSRNRL-----NQYKCDGAECPTDKYGHSI----------V 1131
                                                                    SPTLDSLT-FNNIDGA--------IISGSAYKGSKVDIYNKNGDWLASTIT
Db   805 SRLRLAGGGFDWQRGRYRYDGLGQLVGIDDNQY------PSVRYEYDLGGRLLASRRAGAA 857
                                                                    ESRRYKDKEGH----------DQIVRIERTYNNYHLLTSECKQQNGYI-QTETAYYAI
Qy  1132 TQNFTYDIYGNITACHTTFADGTEDHATFKFANPTDPCQLTEVHHTHPDMPDNIRLKYDK 1191
                                                                    NEEGKFTLQDLSINTNQEVYAVATYNGY-------SSENSSIGLVTEVPAISIT
Db   858 ASTYRYDAAGN-------RLEGVGEYAR-------EDARQAFAENELYRSGFSRSEVRANQ 904
                                                                    IGHNPDSQPSQFQLPKTKTETWRSADNSYRSEITETTPDESG
Qy  1192 AG-------RVINITDNHGNTENFTYDTLGRL-----QNGQGSVYGYDPLNRLVSQKT 1237
                                                                    KDQNGNLQEFNSNVFDSSGITPFSVMALGEV
Db   905 AGEGPARWAGNRVERIAGN-----RYRFDALGNLVERIGADGERLRLAYDGAQRLVHLTR 959
                                                                    NPLTKVIKDKKTQKIISP-------STHWEY-----YPPA----GEVD
Qy  1238 DTLD---CELYYRETMLVNEV-----RNG--EMIRLLRTGETIIAQQRASKVLLTGTDSQ 1287
                                                                    EEGAQIIISTDKDNISGHPQYITADYTPAVFLETPQFDISGETL
Db   960 DYADGTRLEARYRYDALSRRIAKVVLRDGVEQQVRFGWDGDRQCAEAFARELRTTVHEPG 1019
                                                                    NC-PPEPYGFTRFVK---KIIQTPYDSEFKDDPEKFIQYR-YSLIGSQSHVTLKIEERHY
Qy  1288 QSVIL-------TSDKQNLSQ--EAYSAYGKHKSTANDASILGVNGERADPVSGVTHLGN 1338
                                                                    SVHVNEPNSFIRAFSGEGNLIATGP-------TDEQGFASLQVFQFLKEGETVTVQVVDKNQ
Db  1020 GFVPLLRLEQACEPDPPELLQLRQAFAAEGQPL-------PAQCVPALGE 1062
                                                                    SATQLL--------------------NSTLFQYNTDKSELGRLLKQTECTKGE
Qy  1339 GYRSYDPTLMRFHTPDSLSPFGAGGINPYISYCLGDPINRSDPSGHLSWQA---WTGI 1392
                                                                    NTSETLIEVPNFAYIPHVERITQEGLISGVAEDNSTVIVRDADGNELGKV
Db  1063 ARIAF------FHTDH---------LGTPLQLSDERGQLRWQGSVPDDWRAV 1098
                                                                    NGKTYSVVHKFTYTKQDDTLQQSHSITTHDNFTIHRSQVRSRYTG--------R
```

```
Qy   253 IGKNGILGQWIT 292
Db   711 VTQ----BWTEIPABFLWSVEPQTNISIDISLTDQAGNTAIITHTQN--YNVDPTPN 762
Qy   293 LPVLPVTLMKQVPGAGQPAIQAEYSYTSHNYVGGGSNGIWNNKLDNLYGLMTEYNYGST 352
Db   763 SPTLDSLT-FNNIDGA--------IISGSAYKGSKVDIYNKNGDWLASTIT 804
Qy   353 ESRRYKDKEGH----------DQIVRIERTYNNYHLLTSECKQQNGYI-QTETAYYAI 400
Db   805 NEEGKFTLQDLSINTNQEVYAVATYNGY-------SSENSSIGLVTEVPAISIT 851
Qy   401 IGHNPDSQPSQFQLPKTKTETWRSADNSYRSEITETTPDESG 442
Db   852 RISPEGVISGYATEGSHFIV-------KDQNGNLQEFNSNVFDSSGITPFSVMALGEV 903
Qy   443 NPLTKVIKDKKTQKIISP-------STHWEY-----YPPA----GEVD 474
Db   904 RPFILSLDQPL------EEGAQIIISTDKDNISGHPQYITADYTPAVFLETPQFDISGETL 958
Qy   475 NC-PPEPYGFTRFVK---KIIQTPYDSEFKDDPEKFIQYR-YSLIGSQSHVTLKIEERHY 529
Db   959 SVHVNEPNSFIRAFSGEGNLIATGP-------TDEQGFASLQVFQFLKEGETVTVQVVDKNQ 1013
Qy   530 SATQLL--------------------NSTLFQYNTDKSELGRLLKQTECTKGE 562
Db  1014 NTSETLIEVPNFAYIPHVERITQEGLISGVAEDNSTVIVRDADGNELGKV-------- 1063
Qy   563 NGKTYSVVHKFTYTKQDDTLQQSHSITTHDNFTIHRSQVRSRYTG--------R 608
Db  1064 ----------TLGDD---NSWSDFSHFSLSVNRPLIDGEKISVQIIDNKGLMSPEQN 1107
Qy   609 LFSDTDTKDIVTQMSYDKLGRLLTRTLNSGTPYANTLTYD--------YELNNLQDD- 657
Db  1108 IIVDLTPPAPTELNFNDAGDLIV---YGHABPFSEILVKDGQGNILNKWFWNNWTDESGS 1164
Qy   658 -----------NRPPFVITTDVNGN-----------QLRN-------EFDG-----AGRHV 685
Db  1165 FSIELGTFLTLNAETVYVTATDVNGNVSLAAQIQAPNYAFAPYDSFTSDGVISGQAENNS- 1224
Qy   686 SQCLKDSDGDGKFYTIHTCQYDEQGRHHTSTYSYDYLTNGRQQTDPDKVHLSMSKSYDNWG 745
Db  1225 TLVVKDAKGD----VVABIKVGEDNGWNGSSYFK-LQLDRPLVDGEQFFLSIK----DARG 1276
Qy   746 QIANTHWSYGVSEKITVDPITLT-ATKQLQSNSNNVQTG-KEVTTYTPSQQPIQITLFDE 803
Db  1277 QVS------ADTVITADTVAPTPASNLVFSEDGSYLTGVAELNT----------TIQVFDH 1321
Qy   804 AGHLQSC--HTLTRDGWDRVRKETDAIGQCTIY-----------QYDNYNRVIQ 844
Db  1322 NGQLVNIWNNTINSD-----GTFTIYLGSNNLHGEAFTVTVKDQAGNVSEAIS 1369
Qy   845 ITLPDGTIVNRKYAPFSTDTLITDIRVNGISLGQ---------TFDGLSRLTQSQDGGRV 896
Db  1370 INAPLDDI--APNPIKNILLD--ANGQNFTAQAEANSQIEVFDSLG---NQTG-- 1415
Qy   897 WAYTYSAGNDQCPSTVITPDGQFIH-YQYQPELDDAVLQVASNE-----ITQQFSYN 947
Db  1416 WGSTDSAGN-------ASGSFNQTYLHGEELTFVVIDRAGNRSIEFKQNALIDTIAPN 1466
Qy   948 PV-----TGALLKAVAEGQSLTPIYYPSGRLKMENINDMKKMSYLWTLRGLENGYTDL 1000
Db  1467 PIENIIFNENGQSFTAQAERAGSSIDVLDQTG----------NKIGF-------GYTDS 1507
Qy  1001 TGTIQKISRDT--HG---------RVTQIKDSIKTTLNYDDLNRHIGSQVTDLAT 1045
Db  1508 SGNVSGYFQQVYLHGEELTFVVIDRAGNRSAEVKQSA---LNDDVAPNPIENILLDL-N 1562
Qy  1046 GHMLT------TTVEFDGLNRE-IGRKLCDSSGHTLD-IQQSWLKTQQLANRIVKLNGVL 1097
Db  1563 GQNFTAQAEANSQIEIKNNNGDVVGYGSADSAGNVSGYLYQVHLHGEELTFIVVDRAG--- 1620
```

RESULT 15
US-09-328-352-5486
; Sequence 5486, Application US/09328352
; Patent No. 6662958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5486
; LENGTH: 3290
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5486

```
Query Match    2.5%; Score 222.5; DB 4; Length 3290;
Best Local Similarity  18.8%; Pred. No. 5.4e-07;
Matches  370; Conservative 249; Mismatches 656; Indels 697; Gaps 101;

Qy    12 ITMSDNNEF------FTQANNFTSAVSG----------------GVDPRT 39
Db   459 LTLSDLNDTGVSATDQITQDKNFNLKLEGQETGSQVTYLVSTDEGKTWQETTVAQKDLAD 518
Qy    40 GLYNIQITLGHIVGNGN------LGPTLPLT--LSYSPLNKTDIGPGIGFNFGLSVYD 89
Db   519 GIYQYKAVVTDVAGNISETAIQKVVVDTTAPQAGKLITLSDLNDT-----GVSATD 568
Qy    90 ---RKNS-----LLSLSTGE-----NYKVIETDKTVKLQQKKLDNLRFEKDLKENCYRII 136
Db   569 QITQDNSFTLKLVQPIVIGEQAALLDHYEVSKDBGKTWQETTAD----QKDLADGIYQ-- 622
Qy   137 HKSGDIEVLTGFNNNAFDLKVPKLLNPAGHAIYIDWNFEATQPRLNRIYDDLGHDIPL 196
Db   623 YKA-----VVTDLAGNISESAIQKVVDNS-------LNVESTTVIVKPITED------- 663
Qy   197 LNLEYQGLIKTILTLFPPGQKEGYRTELRFPLNRQLNSIHNFSLGNENP----LTWSFGVTP 252
Db   664 NTISLVEKDQVISIRLEIANLPTDLNS----SLTSVNTTLENVTYNFHFDE 710
```

```
Qy  1098  QRTEQYSYDSRNRLNQYK------CDGAECPTDKYGHSIVTQNFT-------------  1136
                            :           :
Db  1621  -------------NRSTEVKQNALIDDIAPNPIENIVLDINGQNFTAQAEANTQIEVKNAVG  1669

Qy  1137  -------YDIYGNITA-CHTTFADGTEDHATFKFAN------PTDPCQLTEVHHTHPDMPD  1183
                   ||::|  :  :           |     |   :::  |   :
Db  1670  EIVGSSGYVDGAGNVSGYLYQVYLHGEE--LTFVVVDRAGNRSTEVKQNALIDDIAPNPIE  1727

Qy  1184  NIRLKYDKAGRVINITDNHGNTENFT----YDTLGRLQNGQGSVYGYDPLNRLVSQKTDT  1239
              ||:|              :||||    :  ::|:|              ::|:
Db  1728  NILL------------DANGQNFTAQAEANTQIEVKNAAGEVIG---------SGSTDS  1765

Qy  1240  L-DCELYYRETMLVNEVRNGEMIRLL---RTGETIIAQQRASKVLLTGTDSQQSVILTSD  1295
             :  :    :    |           :    ::  |::|:::    :
Db  1766  MGNVSGYFYQVYL-----HGEELTFVVVDRAGNRSTEVKQNALIDDIAPNAIENIIFNEN  1820

Qy  1296  KQNLSQEA-YSAYGKHKSTANDASILGYNGERADPVSG-------------VTHLGN  1338
           ||:|:  :   |: :|    :||   :  :     :|||||
Db  1821  GQNFTAQAEANSKVEVKNAAGEVVGSGY-VDSAGNVSGYLNQVYLKGEELTFVVIDQAGN  1879

Qy  1339  GYRSYDPTLMRFHTPDSLSPFGAGGI-----NPYSYCLGDP---INRSDPSGHL------  1384
           |:|                 :  : :     : :|            |||
Db  1880  --RSIEVKQTAF--LDNTAPENATNLVFSEDGSYLSGMAEPNATIQIFDQYGQLLNQWNN  1935

Qy  1385  --SWQAWTGIGMG---IAGLLLTIATGGMAIAAAGGIAAAIASTSTTALAFGALSVTSDI  1439
              :    :         :  :           :: |
Db  1936  NVNWDGTFNIYLNSNYMHGEVFKV----VVVDQAGNLSGEVT-------------------  1973

Qy  1440  TSIVSGALEDASPKASSILGWVSMGMGAAGLAESAIKGGTKLATHIGAFAEDGENALLKS  1499
           :|:|  :|:|         |:|:|   :|||: |  ::|:  :|:    :|:
Db  1974  ---VKAPLDDIAPVAASDLVFNEDGSSLSGVAEP--------NTFIQIFDQNGQQ--MNT  2020

Qy  1500  TSESSRIKWGVTRSLDREIVRNEEGQVIKDHSRGYTDNFMGKGEQAILVHGDKDGFLYHT  1559
           |::|           |  :             ::  |:|    ::|||
Db  2021  WSQSVNADGTFTIFPGTYNLHGEEFTVI------VKDLAGNVSEAVSV-----------  2062

Qy  1560  EGNKHNGKGPYTRHTPEQLVDYLKDNNIVDLTQGGD--KPVHLLSCYGKSSG  1609
           | :|  :||       :|  |||  :: |   :|||      :||::|  |
Db  2063  ------KAPLDDIAPNPIKNIVFDANGQSFTAQAEANSQIEIFDSFGSQIG  2107
```

Search completed: February 9, 2005, 18:20:23
Job time : 74 secs